FONDAMENTI DI INFORMATICA

Massimiliano Salfi

massimiliano.salfi@unict.it

CALENDARIO LEZIONI

In presenza

martedì, 11.00 -13.00 (edificio 14, aula D02)

giovedì, 14.00 - 17.00 (edificio 14, aula D02)

venerdì 8.00 – 10.00 (edificio 14, aula D02)

RIFERIMENTI

Sito internet

http://www.dmi.unict.it/salfi/home.htm

Ricevimento

Solo per appuntamento.

Esami

Progetto e scritto/orale ?!?

PROGRAMMA

- Introduzione all'informatica
- Codifica e rappresentazione delle informazioni
- Algebra booleana
- Architettura dei calcolatori
- Sistema Operativo e applicativi software
- Reti di calcolatori ed internet
- Ipertesti e codice HTML
- La sicurezza in rete ed i malware
- Introduzione alle basi di dati
- Intelligenza Artificiale
- Il linguaggio di programmazione C
- Il pacchetto Microsoft Office

MATERIALE DIDATTICO

- ✓ *Slides del docente*
- ✓ Massimiliano Salfi
  Fondamenti di informatica per ingegneri civili, ambientali e gestionali
  Seconda Edizione – ISBD 9781307789744
  McGraw-Hill

Il testo raccoglie le sole parti che tratteremo, dai seguenti testi:

- ✓ Luca Mari, Giacomo Bonanno, Donatella Sciuto
  Informatica e cultura dell'informazione
  McGraw-Hill - Seconda edizione
- ✓ Alessandro Bellini, Andrea Guidi
  Linguaggio C
  McGraw-Hill - VI edizione
- ✓ Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein
  Introduzione agli algoritmi e strutture dati
  McGraw-Hill - Terza edizione

INTRODUZIONE ALL'INFORMATICA

CHE COS'È L'INFORMATICA?

Informatica

=

Informazione + Automatica

Si riferisce ai processi e alle tecnologie che rendono possibile l'immagazzinamento e l'elaborazione delle informazioni.

GLI ANTENATI DEL COMPUTER

- Macchina analitica di Babbage (1830).
- Macchina di Turing (1936).
- Macchina di Von Neumann (anni ‘40).

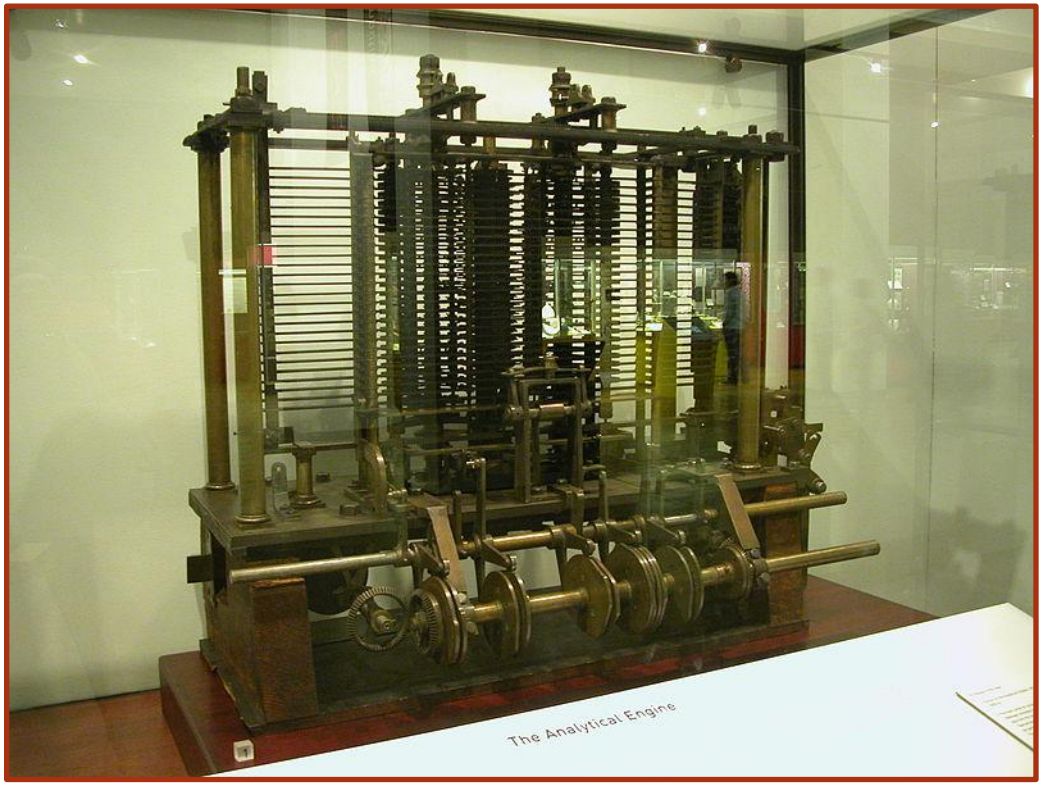
GLI ANTENATI DEL COMPUTER

Il progetto della Macchina analitica di Babbage fu sviluppato dal matematico, filosofo e scienziato inglese Charles Babbage (1791–1871).

Pur essendo realizzata solo in parte per motivi politici e finanziari, rappresenta un importante passo nella storia dell'informatica.

Il moderno PC, infatti, presenta parecchie analogie con il progetto sviluppato da Babbage.

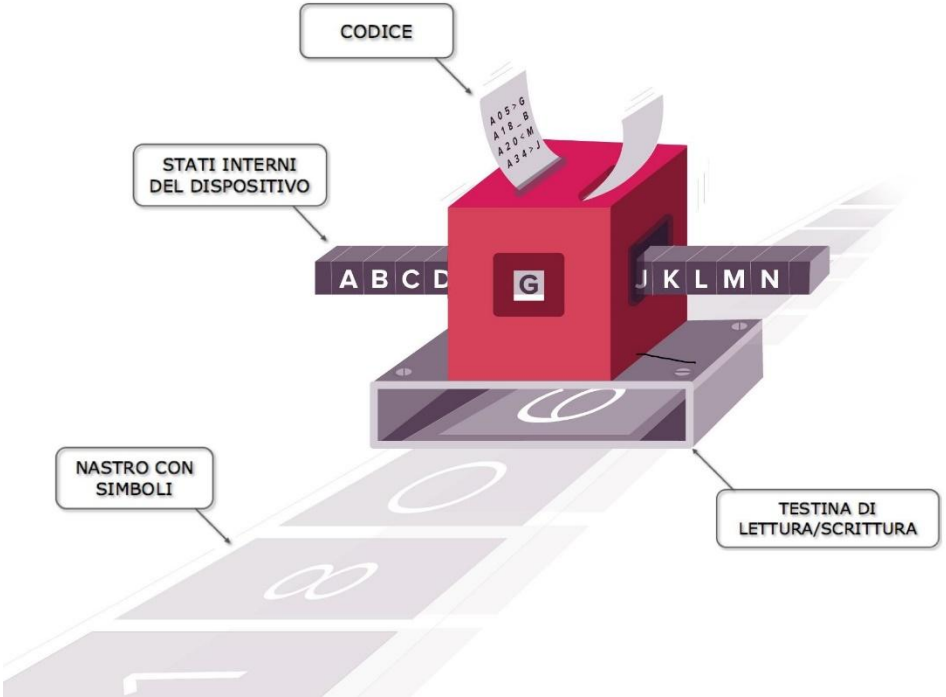
GLI ANTENATI DEL COMPUTER



Modello di una parte dell'Analytical Engine di Babbage in mostra al Museo della scienza di Londra

GLI ANTENATI DEL COMPUTER

La macchina di Turing è una macchina ideale che manipola i dati contenuti su un nastro (memoria) di lunghezza infinita, con cui interagisce attraverso una testina di lettura/scrittura. Il dispositivo è caratterizzato da n (numero finito) stati interni.



GLI ANTENATI DEL COMPUTER

Ogni volta che la testina legge un dato dal nastro, la macchina di Touring può:

- cambiare stato interno (tra quelli ammessi);
- leggere/scrivere un nuovo dato da/sul nastro (memoria);
- spostare la testina, di una cella di memoria, avanti o indietro.

La macchina di Touring può essere programmata attraverso un codice che descrive le azioni che essa deve compiere.

In altre parole, è un modello astratto che definisce una macchina in grado di eseguire algoritmi e dotata di un nastro infinito (la memoria) su cui possono leggere e/o scrivere dei simboli (i dati).

GLI ANTENATI DEL COMPUTER

La macchina di Von Neumann è il modello architetturale secondo il quale è organizzata la maggior parte dei moderni elaboratori.

È dotata delle seguenti parti:

- processore;
- memorie;
- periferiche di Input;
- periferiche di Output;
- sistema di interconnessioni (bus).

Vedremo più avanti questo modello in dettaglio.

LA MACCHINA COMPUTER

In generale, un computer:

- elabora i dati, eseguendo operazioni logiche e aritmetiche;
- salva istruzioni e dati in memoria;
- interagisce con l'ambiente circostante prelevando informazioni da elaborare (input) o fornendo i risultati di tale elaborazione (output).

Un'altra caratteristica importante di un computer è la modularità e la scalabilità.

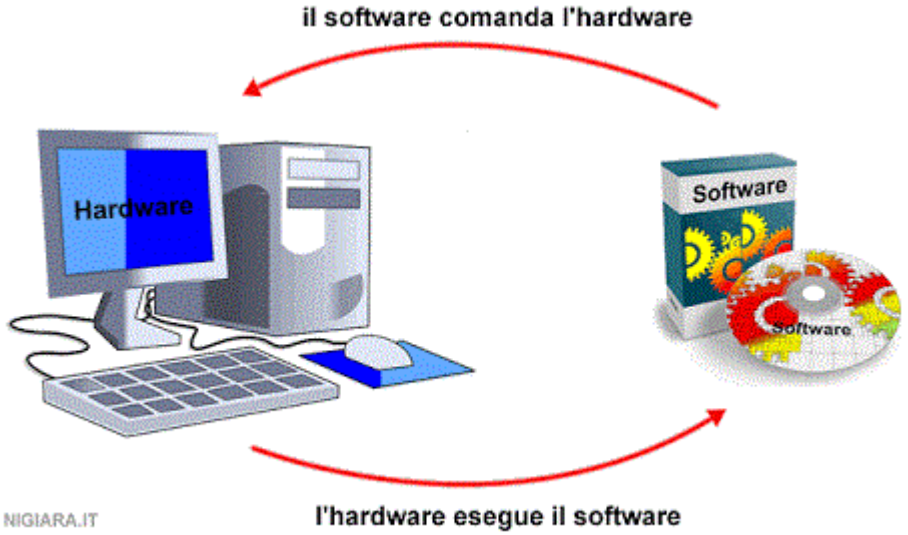
Modulare: utilizzo parti standard, pertanto in caso di guasti posso sostituirli con altri di pari funzionalità.

Scalabile: sostituisco la parti (i moduli) con altre parti compatibili ma aventi caratteristiche migliori.

HARDWARE VS SOFTWARE

L'*hardware* denota la struttura fisica del computer, costituita di norma da componenti elettronici che svolgono specifiche funzioni nel trattamento dell'informazione.

Il *software* indica l'insieme delle istruzioni che consentono alle varie parti hardware di svolgere i compiti per i quali sono stati progettati.



POSSIBILI SCENARI APPLICATIVI

- economico e commerciale;
- industriale;
- didattico e della formazione professionale;
- arte e spettacolo;
- ingegneristico;
- matematico e delle scienze;
- lavorativo e del tempo libero;
- medico;
- etc.

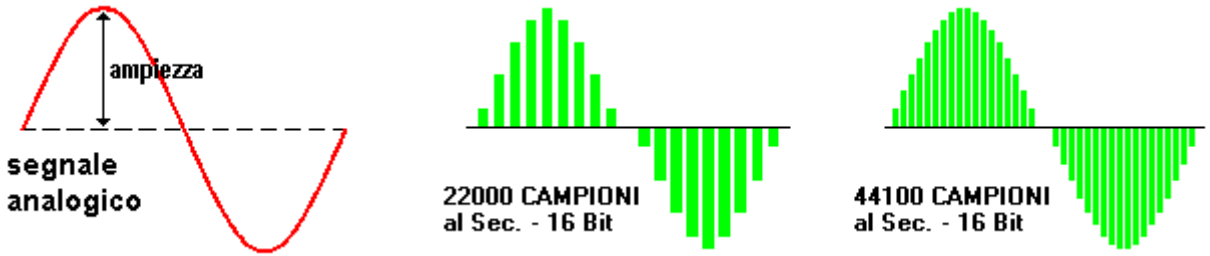
LA CODIFICA E LA RAPPRESENTAZIONE DELLE INFORMAZIONI

I SEGNALI PER COMUNICARE

Gli essere viventi ed il computer utilizzano modi tra loro differenti per comunicare:

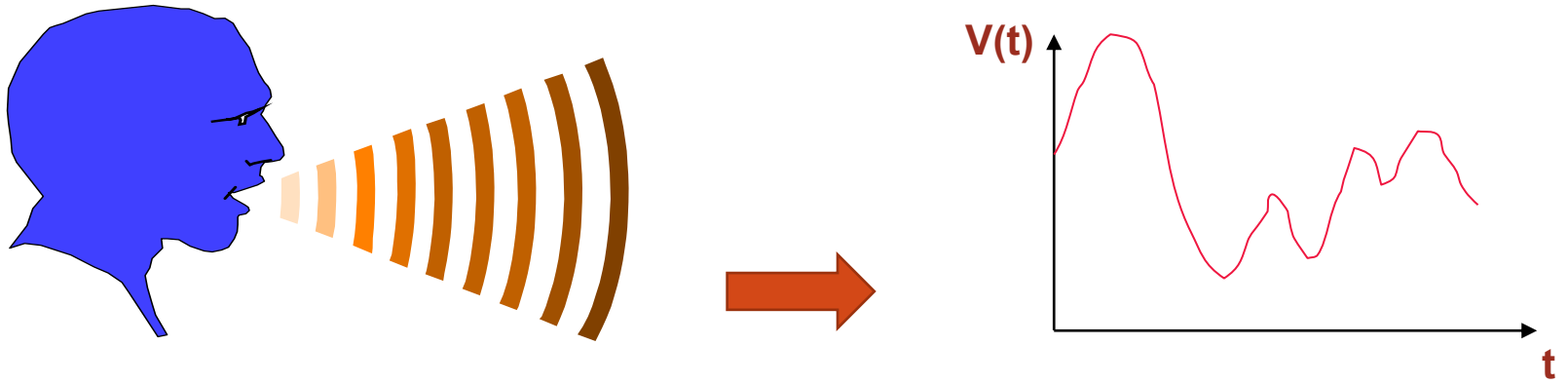
- segnali *analogici*;
- segnali *digitali*.

Ad un segnale analogico faremo corrispondere una grandezza continua, mentre ad un segnale digitale, una grandezza discreta.



INFORMAZIONE ANALOGICA

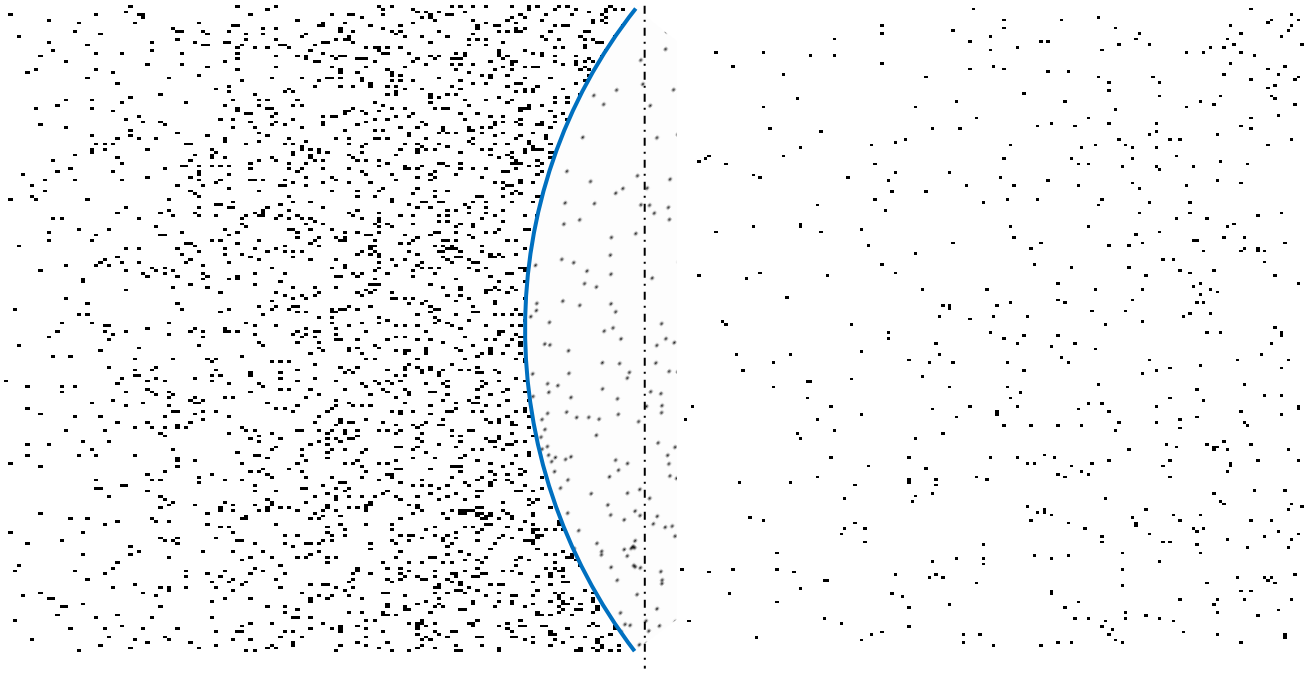
Ad esempio, la voce umana, i suoni presenti in natura o emessi da strumenti musicali, sono sistemi di comunicazione di tipo *analogico*, nei quali le grandezze fisiche che entrano in gioco sono funzioni continue del tempo.



Vediamo in dettaglio cosa accade, ad esempio, quando con il microfono di un computer, durante una call, occorre trasmettere ad un computer remoto la nostra voce, o un qualsiasi suono.

INFORMAZIONE ANALOGICA

Consideriamo, ad esempio, la corda di uno strumento musicale che vibra in un mezzo quale l'aria.



INFORMAZIONE ANALOGICA

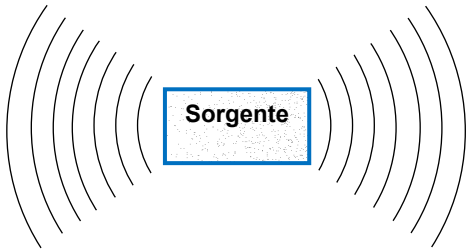
Inizialmente la densità delle molecole d'aria è costante, cioè in ogni unità di volume vi è lo stesso numero di molecole.

Quando viene sollecitata, la corda si sposta avanti e indietro, rispetto alla posizione di equilibrio indicata dalla linea tratteggiata in figura precedente e, durante questo spostamento, comprime le molecole d'aria da un lato e le espande dall'altro.

Ogni volta che il moto si inverte, le molecole che prima erano state compresse si espandono e viceversa.

INFORMAZIONE ANALOGICA

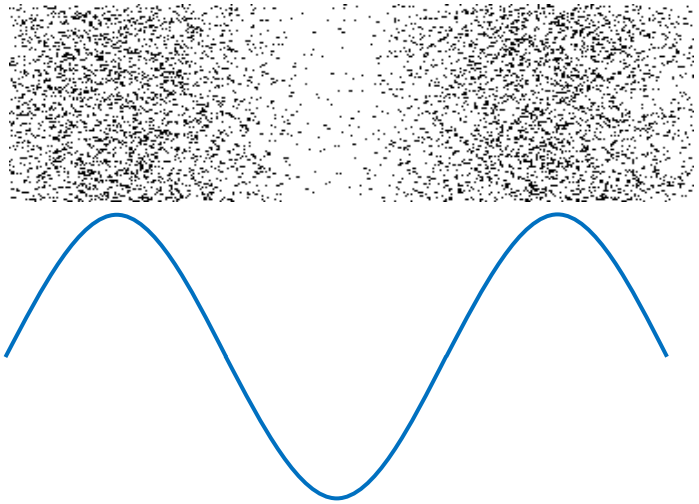
Le compressioni e le espansioni (cioè le perturbazioni dell'aria che inizialmente era in stato di quiete) si propagano poi con una certa velocità attraverso l'aria circostante in tutte le direzioni, dando luogo a onde sferiche.



Queste onde arriveranno fino alla membrana del microfono che convertirà, attraverso un opportuno trasduttore, la variazione di pressione delle particelle di aria, in un segnale elettrico.

INFORMAZIONE ANALOGICA

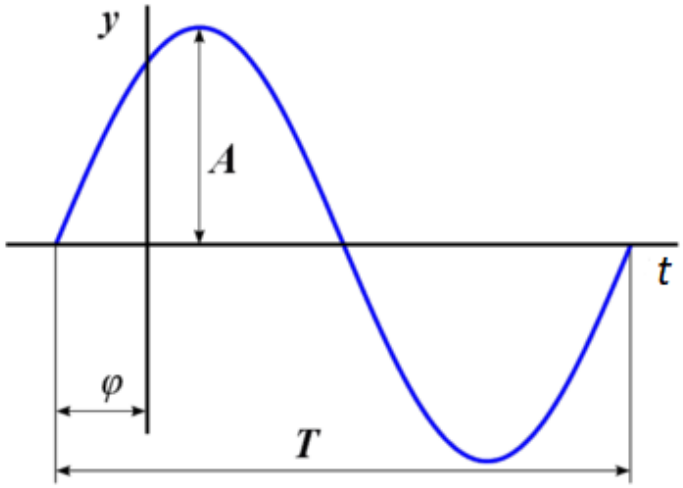
Per quanto riguarda la forma d'onda del fenomeno appena illustrato, si noti che, ad esempio, può essere osservato **dal punto di vista del tempo,** misurando come come varia il valore della pressione, in uno stesso punto (una molecola del mezzo), allo scorrere del tempo:



Osservando il fenomeno dal punto di vista dello spazio (propagazione dell'onda), comunque, la forma d'onda sarà sempre simile.

INFORMAZIONE ANALOGICA

Graficamente otteniamo, quindi, una sinusoide che è una grandezza continua:



$$y = A\sin(2\pi f t + \varphi)$$

Tuttavia, y assume infiniti valori, mentre in un computer, per quanto grande possa essere la memoria di cui è dotato, non potremo acquisire (e trasmettere) tutti valori.

INFORMAZIONE DIGITALE

La codifica dei segnali nei computer avviene in modo digitale, in quanto le grandezze fisiche sono rappresentate da coppie di stati discreti:

(0, 1) oppure (off, on) oppure (false, true)

Nei circuiti digitali il simbolo 0 (oppure *off*, oppure *false*) è associato ad un segnale a basso voltaggio (interruttore spento); di contro il simbolo 1 (oppure *on*, oppure *true*) è associato ad un segnale ad alto voltaggio (interruttore acceso).

DIGITALIZZAZIONE DEI SEGNALI

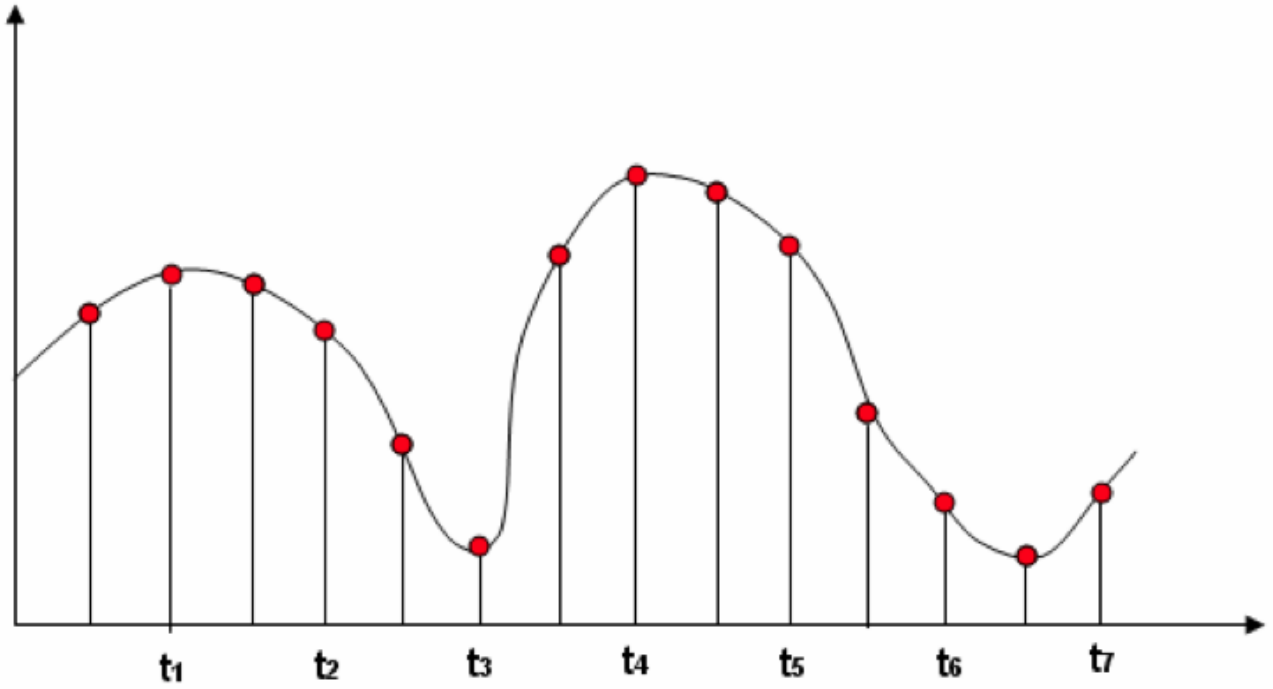
Qualsiasi segnale «reale», dunque, per poter essere elaborato attraverso un computer, deve essere convertito in un segnale digitale (e quindi in una sequenza finita di valori numerici).

Nel farlo, devo attivare un procedimento noto come «campionamento», che mi permette di prelevare un certo numero di valori nell'unità di tempo (frequenza di campionamento).

Tale conversione, per quanto accurata, comporta sempre un certo grado di **approssimazione**, con conseguente perdita di informazione. Inoltre, gli strumenti di misura utilizzati, per quanto precisi, non mi permetteranno mai di disporre di tutti i valori, introducendo un **errore detto di quantizzazione**.

DIGITALIZZAZIONE DEI SEGNALI

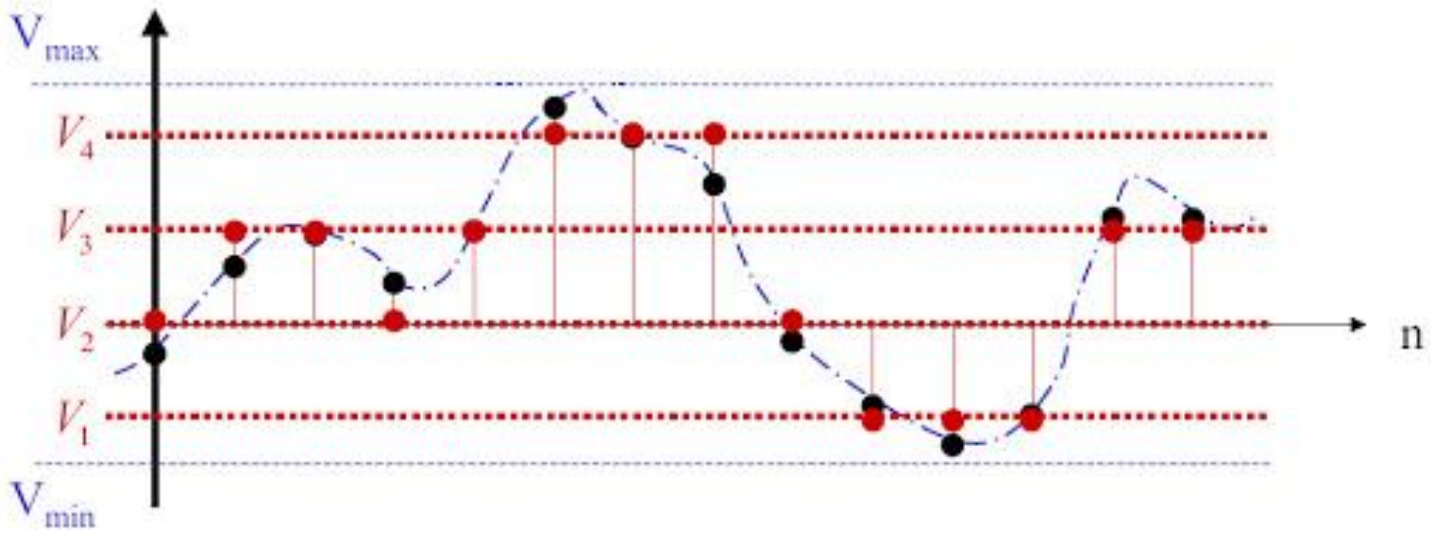
Consideriamo ad esempio una grandezza continua e supponiamo di prelevare 7 campioni. Graficamente avremo:



In questo caso, però, stiamo supponendo di poter misurare esattamente il valore di ciascun singolo campione.

DIGITALIZZAZIONE DEI SEGNALI

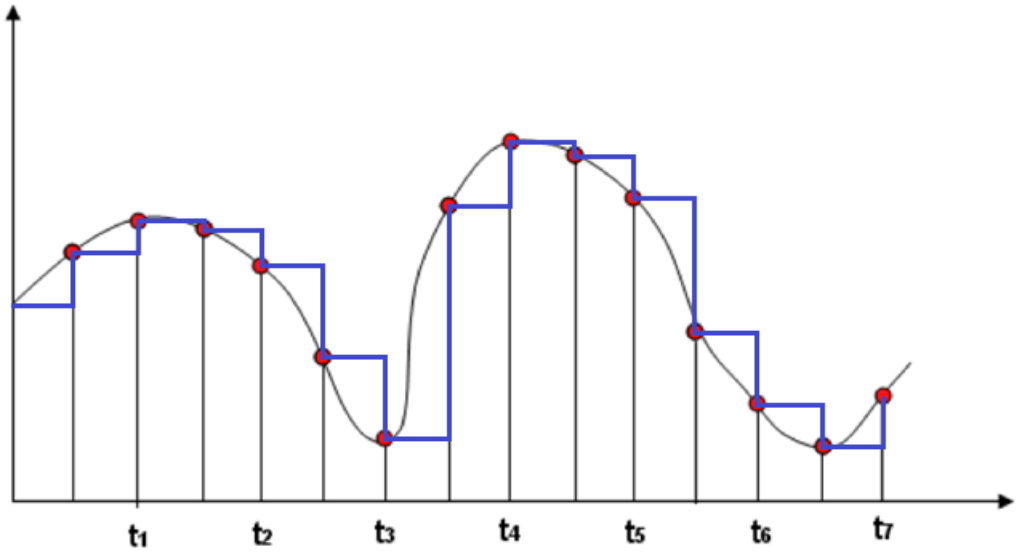
Nella realtà, come detto, dovremo fare i conti con l'errore di quantizzazione, in quanto ciascuna misura risentirà della scala di valori disponibile in ordinata tra V_{min} e V_{max}:



DIGITALIZZAZIONE DEI SEGNALI

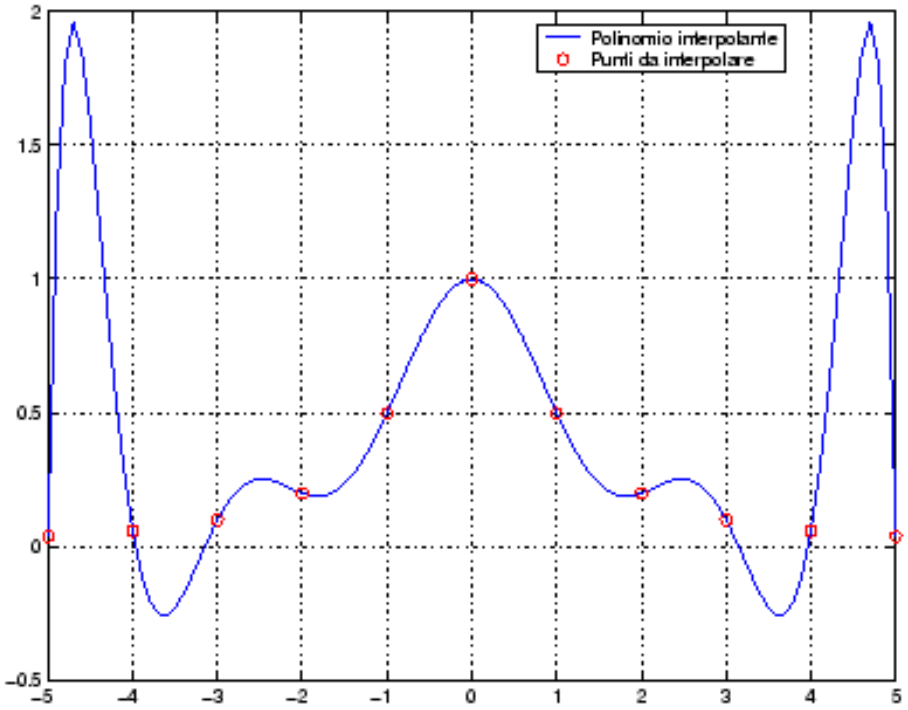
Una volta ricevuto il segnale digitale, il computer remoto dovrà ricostruire il segnale mancante tra i vari campioni. Per la ricostruzione, si possono utilizzare varie tecniche di interpolazione.

Ad esempio, quella del «sample and hold», che mantiene costante il valore fino al punto successivo:



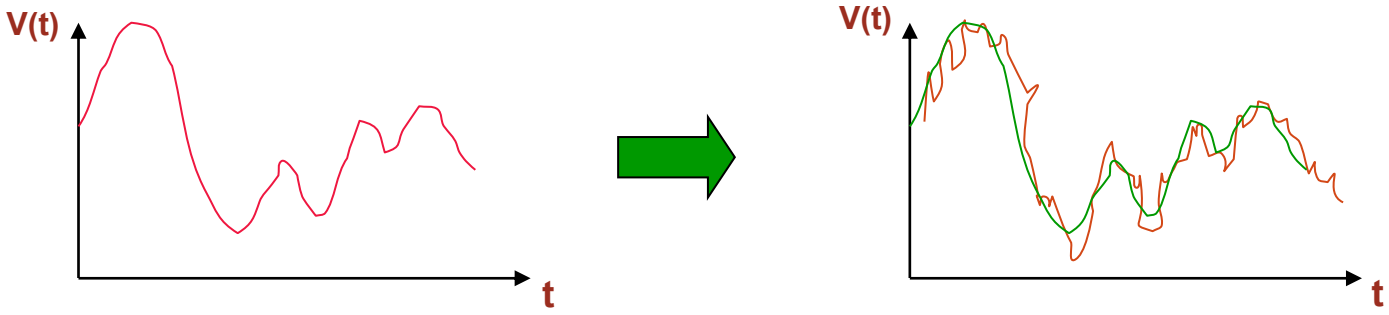
DIGITALIZZAZIONE DEI SEGNALI

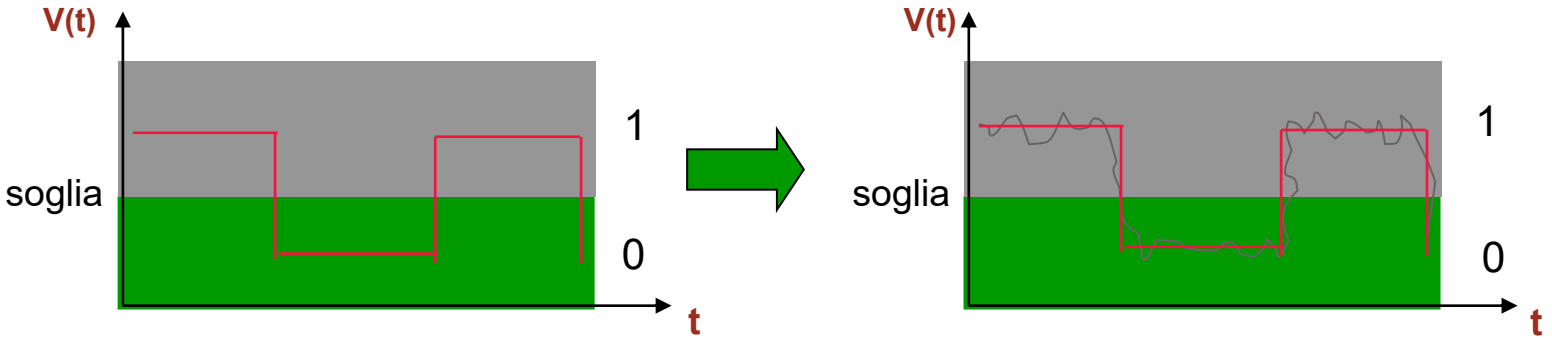
O, per evitare un segnale troppo «spigoloso», attraverso interpolazione polinomiale (ad esempio interpolazione di Langrange):



PRECISIONE DEI SEGNALI

I segnali digitali presentano il vantaggio di essere meno affetti da disturbi di trasmissione (errori) rispetto a quelli analogici.

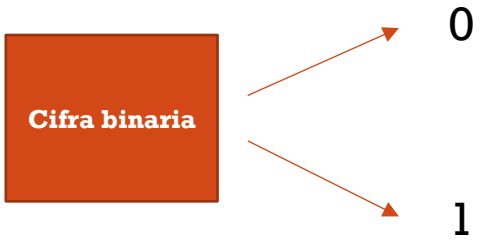




RAPPRESENTAZIONE DELLE INFORMAZIONI

In un calcolatore, tutte le informazioni sono rappresentate in forma binaria (o digitale), come detto, utilizzando due soli simboli. Di solito si adoperano i simboli 0 ed 1.

Con una cifra binaria si possono, quindi, rappresentare soltanto due informazioni.



RAPPRESENTAZIONE DELLE INFORMAZIONI

È semplice intuire, in ragione di quanto detto in precedenza, che tale scelta è legata alla necessità di rappresentare, ad esempio:

- due possibili stati di polarizzazione di una sostanza magnetizzabile;
- il passaggio (o non passaggio) di corrente attraverso un conduttore;
- Il passaggio (o non passaggio) della luce attraverso una fibra ottica.

Questo concetto verrà ripreso in seguito, nella macchina di Von Neumann, quando parleremo di memorie.

IL BIT

Il bit è *l'unità di misura dell'informazione* ed è definita come la quantità minima di informazione che serve a rappresentare due differenti stati: 0 oppure 1. Il nome proviene da **Bi**nary Digi**t**.

Si utilizzano i seguenti multipli :

- Kilo (Kb), pari a 2^{10} ~ un migliaio (1024 bit);
- Mega (Mb) 2^{20} ~ un milione (1024 x 1024 bit);
- Giga (Gb) 2^{30} ~ un miliardo (1Mb x 1024 bit);
- Tera (Tb) 2^{40} ~ mille miliardi (1Gb x 1024 bit).

CODIFICA BINARIA

Per poter rappresentare un numero maggiore di dati, o informazioni, è necessario utilizzare sequenze di bit. Utilizzando stringhe lunghe due bit si possono rappresentare quattro informazioni diverse, vale a dire:

00 01 10 11

In generale, con n bit si possono avere 2^n stringhe (dette anche parole in linguaggio macchina) differenti.

Il processo che fa corrispondere ad una informazione una configurazione di bit, prende il nome di *codifica dell'informazione*.

SEQUENZE DI BIT

Numero di bit nella sequenza	Informazioni rappresentabili
2	4
3	8
4	16
5	32
6	64
7	128
8	256

SET DI CARATTERI

Nella comunicazione scritta, di norma, è necessario disporre dei seguenti caratteri:

- 52 lettere alfabetiche maiuscole e minuscole;
- 10 cifre (0, 1, 2, …, 9);
- Segni di punteggiatura (, . ; : ! ” ? ’ ^ \ …);
- Segni matematici (+, - , × , : , ± , { , [, >, …);
- Caratteri nazionali (à, è, ì, ò, ù, ç, ñ, ö, …);
- Altri segni grafici (©, ←, ↑, @, €, …).

CODICE

Si pone quindi la necessità di codificare almeno 220 caratteri utilizzando sequenze di bit.

La sequenza di bit necessaria a rappresentare 220 simboli deve essere composta da 8 bit (con 8 bit, infatti, dispongo di 2^8=256 stringhe differenti).

La corrispondenza tra sequenze (o stringhe) di bit e simboli prende il nome di *codice*.

IL BYTE

Un gruppo di 8 bit viene denominato Byte. Un byte:

- corrisponde ad un carattere;
- rappresenta l'unità di misura della capacità di memoria.

Si utilizzano i seguenti multipli del Byte:

- Kilo (KB) 2^{10} ~ un migliaio (1024 byte)
- Mega (MB) 2^{20} ~ un milione (1024 x 1024 byte)
- Giga (GB) 2^{30} ~ un miliardo (1MB x 1024 byte)
- Tera (TB) 2^{40} ~ mille miliardi (1GB x 1024 byte)

CODIFICA SET DI CARATTERI

Codifiche standard:

- ASCII, 8 bit per carattere, rappresenta 256 caratteri.
- UNICODE, 16 bit per carattere (estende il codice ASCII con i caratteri etnici).

Codifiche proprietarie:

- MSWindows, 16 bit per carattere (molto simile ad UNICODE).

CODICE ASCII

Acronimo di:

American

Standard

Code for

Information

Interchange

Binary	Oct	Dec	Hex	Glyph
010 0000	040	32	20	
010 0001	041	33	21	!
010 0010	042	34	22	"
010 0011	043	35	23	#
010 0100	044	36	24	$
010 0101	045	37	25	%
010 0110	046	38	26	&
010 0111	047	39	27	'
010 1000	050	40	28	(
010 1001	051	41	29	)
010 1010	052	42	2A	*
010 1011	053	43	2B	+
010 1100	054	44	2C	,
010 1101	055	45	2D	-
010 1110	056	46	2E	.
010 1111	057	47	2F	/
011 0000	060	48	30	0
011 0001	061	49	31	1
011 0010	062	50	32	2
011 0011	063	51	33	3
011 0100	064	52	34	4
011 0101	065	53	35	5
011 0110	066	54	36	6
011 0111	067	55	37	7
011 1000	070	56	38	8
011 1001	071	57	39	9
011 1010	072	58	3A	:
011 1011	073	59	3B	;
011 1100	074	60	3C	<
011 1101	075	61	3D	=
011 1110	076	62	3E	>
011 1111	077	63	3F	?

Binary	Oct	Dec	Hex	Glyph
100 0000	100	64	40	@
100 0001	101	65	41	A
100 0010	102	66	42	B
100 0011	103	67	43	C
100 0100	104	68	44	D
100 0101	105	69	45	E
100 0110	106	70	46	F
100 0111	107	71	47	G
100 1000	110	72	48	H
100 1001	111	73	49	I
100 1010	112	74	4A	J
100 1011	113	75	4B	K
100 1100	114	76	4C	L
100 1101	115	77	4D	M
100 1110	116	78	4E	N
100 1111	117	79	4F	O
101 0000	120	80	50	P
101 0001	121	81	51	Q
101 0010	122	82	52	R
101 0011	123	83	53	S
101 0100	124	84	54	T
101 0101	125	85	55	U
101 0110	126	86	56	V
101 0111	127	87	57	W
101 1000	130	88	58	X
101 1001	131	89	59	Y
101 1010	132	90	5A	Z
101 1011	133	91	5B	[
101 1100	134	92	5C	\
101 1101	135	93	5D	]
101 1110	136	94	5E	^
101 1111	137	95	5F	_

Binary	Oct	Dec	Hex	Glyph
110 0000	140	96	60	`
110 0001	141	97	61	a
110 0010	142	98	62	b
110 0011	143	99	63	c
110 0100	144	100	64	d
110 0101	145	101	65	e
110 0110	146	102	66	f
110 0111	147	103	67	g
110 1000	150	104	68	h
110 1001	151	105	69	i
110 1010	152	106	6A	j
110 1011	153	107	6B	k
110 1100	154	108	6C	l
110 1101	155	109	6D	m
110 1110	156	110	6E	n
110 1111	157	111	6F	o
111 0000	160	112	70	p
111 0001	161	113	71	q
111 0010	162	114	72	r
111 0011	163	115	73	s
111 0100	164	116	74	t
111 0101	165	117	75	u
111 0110	166	118	76	v
111 0111	167	119	77	w
111 1000	170	120	78	x
111 1001	171	121	79	y
111 1010	172	122	7A	z
111 1011	173	123	7B	{
111 1100	174	124	7C	|
111 1101	175	125	7D	}
111 1110	176	126	7E	~

CODICE ASCII: ALTRO ESEMPIO

La parola *Computer* in codice ASCII diventa, dunque:

C	=	01000011
o	=	01101111
m	=	01101101
p	=	01110000
u	=	01110101
t	=	01110100
e	=	01100101
r	=	01110010

01000011 01101111 01101101 11100000 01110101 01110100 01100101 01110010

CODICE ASCII

Data una sequenza di bit, dividendola in gruppi di byte è possibile risalire ai caratteri originali:

01000001 01101101 01101111 01110010 01100101 00100001

01000001 01101101 01101111 01110010 01100101 00100001

A m o r e !

CODICE ASCII

Con il codice ASCII è possibile rappresentare i numeri come sequenza di caratteri. Ad esempio il numero 234 sarà rappresentato come:

00110010 00110011 00110100

2 **3** **4**

Con questo tipo di rappresentazione, però, non è possibile effettuare operazioni aritmetiche.

IL SISTEMA DECIMALE

Il sistema di numerazione decimale è un sistema *posizionale*. Questo significa che ogni cifra del numero, assume un valore differente in funzione della posizione. Consideriamo il numero seguente in *notazione compatta*:

221

corrisponde a:

$$\mathbf{2} \times 100 + \mathbf{2} \times 10 + \mathbf{1} \times 1$$

Che, in *notazione esplicita*, si scriverà:

$$\mathbf{2} \times 10^2 + \mathbf{2} \times 10^1 + \mathbf{1} \times 10^0$$

IL SISTEMA DECIMALE

Ogni numero si esprime come la somma dei prodotti di ciascuna cifra per la base, elevata all'esponente dato dalla posizione della cifra:

$$221 = 2 \times 10^2 + 2 \times 10^1 + 1 \times 10^0$$

È quindi basato sulle potenze del 10 ed utilizza, come detto, una notazione posizionale basata sui 10 simboli:

$$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

LA NOTAZIONE POSIZIONALE

La notazione posizionale può essere usata con qualunque base, creando così differenti sistemi di numerazione. Per ogni base di numerazione si utilizza un numero di cifre uguale alla base.

In informatica si utilizzano prevalentemente i sistemi di numerazione:

- binaria,
- ottale,
- esadecimale.

Il sistema di numerazione romano non è posizionale:

XIII vs. CXII

IL SISTEMA BINARIO

Il sistema di numerazione binario utilizza una notazione posizionale basata sulle 2 cifre:

$$\{0, 1\}$$

e sulle potenze di 2. Il numero binario 1001 può essere rappresentato esplicitamente come:

$$1001_2 = 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 9_{10}$$

IL SISTEMA OTTALE

Il sistema di numerazione ottale utilizza una notazione posizionale basata sulle 8 cifre:

$$\{0, 1, 2, 3, 4, 5, 6, 7\}$$

e sulle potenze di 8. Il numero ottale 534 può essere rappresentato, esplicitamente, come:

$$534_8 = 5 \times 8^2 + 3 \times 8^1 + 4 \times 8^0 = 348_{10}$$

IL SISTEMA ESADECIMALE

La numerazione esadecimale utilizza una notazione posizionale basata sulle 16 cifre

{0 ,1 ,2 ,3 ,4 ,5 ,6 ,7 ,8 ,9, A, B, C, D, E, F}

e sulle potenze di 16. Si noti che A=10, B=11, C=12, D=13, E=14 ed F=15. Il numero esadecimale $B7FC_{16}$ può essere rappresentato esplicitamente come:

$$(11) \times 16^3 + 7 \times 16^2 + (15) \times 16^1 + (12) \times 16^0 = 47100_{10}$$

CONVERSIONE DA BASE B A BASE 10

Per convertire un numero da una qualunque base b alla base 10, è sufficiente rappresentarlo esplicitamente:

$$1101_2 = 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 13_{10}$$

$$710_8 = 7 \cdot 8^2 + 1 \cdot 8^1 + 0 \cdot 8^0 = 456_{10}$$

$$A51_{16} = (10) \cdot 16^2 + 5 \cdot 16^1 + 1 \cdot 16^0 = 2641_{10}$$

....

$$C_{n-1}C_{n-2}\dots C_1\ C_0 = C_{n-1}b^{n-1} + C_{n-2}b^{n-2} + \cdots + C_1 b^1 C_0 b^0$$

dove ogni cifra $C_i \epsilon [0, b[$ per $i = 0,1, \dots n-1$ (ad esempio, per b=8 ogni cifra, in decimale, assume valore appartenente a [0, 7]).

CONVERSIONE DA BASE 10 A BASE N

Per convertire un numero decimale ad una base n qualsiasi, occorre trovare tutti i resti delle successive divisioni del numero per la base n.

Proviamo, ad esempio, a trovare il valore binario del numero 210. Come mostrato nella slide successiva, basterà dividere 210 per la base 2, ripetutamente, fino quando non si ottiene zero come risultato e leggere i resti dall'ultimo verso il primo.

CONVERSIONE DA BASE 10 A BASE 2

210	2	resto	0
105	2		1
52	2		0
26	2		0
13	2		1
6	2		0
3	2		1
1	2		1
0			

Leggendo la sequenza dei resti dal basso verso l'alto, si ottiene il numero:

11010010_2

VERIFICA DI CORRETTEZZA

Per una verifica di correttezza basta riconvertire il risultato alla base 10:

$$11010010_2 = 1 \times 2^7 + 1 \times 2^6 + 0 \times 2^5 + 1 \times 2^4 +$$

$$+ 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 210_{10}$$

COSTRUZIONE DEI NUMERI BINARI

Per poter costruire la successione dei numeri binari, si può seguire il seguente schema:

0	0	0	0	=	0
0	0	0	1	=	1
0	0	1	0	=	2
0	0	1	1	=	3
0	1	0	0	=	4
0	1	0	1	=	5
0	1	1	0	=	6
0	1	1	1	=	7

RAPPRESENTAZIONE DEI NUMERI

All'interno dei computer, a causa di vincoli tecnologici, per rappresentare qualsiasi tipo di numero, si utilizza sempre un numero fisso di cifre binarie.

I valori più utilizzati sono:

- 16 bit (2 byte)
- 32 bit (4 byte)

In alcuni casi si può arrivare anche a 64 bit (8 byte) o più a seconda del tipo di processore (vedremo in dettaglio tutto questo, quando parleremo della macchina di Von Neumann).

RAPPRESENTAZIONE DEI NUMERI

Tutti i numeri vengono distinti in tre categorie:

- Interi senza segno (interi positivi).
- Interi con segno (interi positivi e negativi).
- Reali (numeri positivi e negativi con virgola).

Ogni categoria viene rappresentata in modo differente.

RAPPRESENTAZIONE CON NUMERO FISSO DI CIFRE

Per comprendere il meccanismo alla base della rappresentazione con un numero fisso di cifre partiamo da un interrogativo:

Qual è il numero più grande rappresentabile con 4 cifre?

- In base 10, avremo 9999
- In base 2 avremo 1111 = 15_{10}
- In base 8 avremo 7777 = 4095_{10}
- n base 16 avremo FFFF= 65535_{10}

RAPPRESENTAZIONE CON NUMERO FISSO DI CIFRE

In generale, con n cifre vale la relazione seguente:

$$b^n - 1$$

essendo b la base del sistema di numerazione. Infatti:

- In base 10, avremo $9999 = 10^4 - 1$
- In base 2 avremo 1111 $15_{10} = 2^4 - 1$
- In base 8 avremo 7777 $4095_{10} = 8^4 - 1$
- n base 16 avremo FFFF $65535_{10} = 16^4 - 1$

RAPPRESENTAZIONE CON NUMERO FISSO DI CIFRE

Vale quindi la seguente regola:

Nella base di numerazione b, disponendo di n cifre, si possono rappresentare soltanto i numeri da

0 a b^n-1

RAPPRESENTAZIONE DEI NUMERI INTERI SENZA SEGNO

In un elaboratore, tenuto conto di quanto detto, considerato che nelle rappresentazioni si adotta il sistema binario e che le parole sono lunghe 16 o 32 bit, si ha:

- Nella rappresentazione a 16 bit i possibili valori saranno compresi tra 0 e 65.535;
- Nella rappresentazione a 32 bit i possibili valori saranno compresi tra 0 e 4.294.967.295

RAPPRESENTAZIONE DEI NUMERI INTERI CON SEGNO

Per rappresentare i numeri con il loro segno (interi positivi e negativi) esistono due possibili modi:

- Rappresentazione in modulo e segno;
- Metodo del complemento a due.

Nel primo metodo, dati n bit, il bit a sinistra si riserva al segno, lasciando gli altri $n-1$ alla rappresentazione del numero. Di solito si considera 0 per il segno più (+) e 1 per il meno (-):

$$0000\ 0101_2 = +\ 5_{10}$$

$$1000\ 0101_2 = -\ 5_{10}$$

RAPPRESENTAZIONE DEI NUMERI INTERI CON SEGNO

Questo metodo, anche se molto semplice, presenta l'inconveniente che esistono due zeri:

1 0 0 0 0 0 1 1	= – 3
1 0 0 0 0 0 1 0	= – 2
1 0 0 0 0 0 0 1	= – 1
1 0 0 0 0 0 0 0	= – 0
0 0 0 0 0 0 0 0	= + 0
0 0 0 0 0 0 0 1	= + 1
0 0 0 0 0 0 1 0	= + 2
0 0 0 0 0 0 1 1	= + 3

RAPPRESENTAZIONE DEI NUMERI INTERI CON SEGNO

Utilizzando n bit e riservandone uno al segno, l'applicazione della formula precedente porterà alla rappresentazione del seguente intervallo:

$$\text{da } -(2^{n-1} - 1) \text{ a } 2^{n-1} - 1$$

dove, come detto, n vale o 16, permettendo quindi la rappresentazione dei numeri da -32.767 a +32.767, o 32 con rappresentazione dei numeri da -2.147.483.647 a +2.147.483.647

RAPPRESENTAZIONE DEI NUMERI INTERI CON SEGNO

Il secondo metodo è quello del complemento a 2:

> Dato un numero composto da n bit, la rappresentazione in complemento a due si ottiene invertendo gli 1 in 0 e gli 0 in 1 (complemento a 1) e poi sommando 1 al risultato ottenuto.

In questo caso, ad ogni numero (senza segno) vengono aggiunti degli zeri a sinistra, fino a raggiungere un byte (o multipli di byte). Se il numero è positivo, si lascia invariato, se è negativo si converte in complemento a due (ottenendo sempre il primo bit = 1, come in precedenza).

È ovvio che 00000000 e 10000000 hanno significati differenti.

OPERAZIONI CON I BINARI

Per procedere con la somma tra binari, come operazioni in colonna, occorre tenere conto delle seguenti indicazioni:

riporto

			1
0 +	0 +	1 +	1 +
0 =	1 =	0 =	1 =
0	1	1	0

Dal momento che un computer è in grado di compiere solo somme e confronti, ci serve una procedura che consenta di convertire ogni operazione, in somma (o sequenze di somme).

Vediamo come farlo in decimale (il procedimento può essere esteso, poi, ai binari con lo stesso ragionamento).

OPERAZIONI CON I BINARI

SOTTRAZIONI
La sottrazione può essere convertita in somma nel modo seguente:

$$25 - 14 = 25 + (-14)$$

MOLTIPLICAZIONI
La moltiplicazione può essere convertita sommando un fattore, per sé stesso, tante volte quanto è il valore assunto dall'altro fattore (in genere si sceglie il minimo tra i due):

$$25 \times 3 = 25 + 25 + 25$$

DIVISIONE
La divisione può essere ottenuta sottraendo il divisore al dividendo e iterando il procedimento fino a quando il risultato non diventi minore dello stesso divisore:

$$10 : 3 \rightarrow 10 - 3 = 7 - 3 = 4 - 3 = 1$$

Il numero di sottrazioni effettuate corrisponde al risultato mentre il valore finale ottenuto, indica il resto della divisione.

L'OVERFLOW

Come visto, per questioni tecnologiche tutti i computer, senza alcuna eccezione, trattano i numeri sempre con una quantità fissa di cifre binarie (ad esempio 16, 32 o più).

Quando l'elaboratore esegue un'operazione il cui risultato eccede il numero di cifre permesso, la computazione si arresta immediatamente e viene segnalato l'errore di OVERFLOW.

L'OVERFLOW

Ad esempio, se la rappresentazione è a 32 bit senza segno e si vuole eseguire la seguente operazione:

3.000.000.000 x
2.000.000.000 =

********** Errore di **OVERFLOW**

In questo caso l'elaborazione si arresta.

ALGEBRA DI BOOLE

INTRODUZIONE

L'algebra booleana deve il proprio nome al matematico inglese George Boole che la formalizzò nel 1854.

Fu sviluppata per poter trattare qualsiasi tipo di proposizione, ovvero di espressione le cui variabili possono assumere solo due valori:

True/Vero False/Falso

Deve la sua importanza al fatto che i circuiti elettronici digitali possono essere rappresentati con una funzione booleana e viceversa.

INTRODUZIONE

Un circuito elettronico digitale, infatti, è un sistema caratterizzato dalla presenza di segnali che possono assumere due soli valori di tensione:

0V +5V

(anche se, in alcune tipologie di circuiti, si preferisce parlare di Low, o L, e High , o H, al fine di considerare valori differenti).

In genere si fa corrispondere il simbolo 0 al valore di tensione 0V (o Low, o L) ed il simbolo 1 al valore di tensione +5V (o High, o H).

L'algebra di Boole è diventata anche l'elemento fondamentale per la logica formale che costituisce la base per il ragionamento scientifico.

INTRODUZIONE

Secondo Boole, dunque, il ragionamento è basato su proposizioni che possono assumere valore Vero o Falso. Ad esempio:

Oggi piove!

è una proposizione che può essere vera o falsa. Le proposizioni possono anche essere combinate tra loro. Ad esempio, consideriamo:

«Piove» «Esco» «Apro l'ombrello»

Tutte e tre possono assumere valore Vero o Falso. Secondo Boole, queste possono anche essere combinate tra loro. Ad esempio:

Se «Piove» e «Esco» allora «Apro l'ombrello»

In questo caso, se sono vere le prime due, allora è vera anche l'ultima.

INTRODUZIONE

L'algebra di Boole, dunque, introduce i concetti di **Variabili** e **Funzioni**.

Una *variabile booleana* viene rappresentata, solitamente, attraverso una lettera maiuscola (A, B, C,…) e può assumere, come le proposizioni, solo valore *Vero* o *Falso*.

Una *funzione booleana*, invece, è una legge che ammette in input n variabili booleane e che fornisce in output un valore che può essere, ancora una volta, solo *Vero* o *Falso*.

Il comportamento di una funzione booleana può essere descritto attraverso la **tavola di verità**, ovvero una tabella avente 2^n righe (se n è il numero di variabili in input) che ci permettono di verificare per quali combinazioni di ingresso è vera la funzione.

INTRODUZIONE

ESEMPIO

Consideriamo la funzione:

$$Y = f(A, B, C)$$

La tabella a lato indica, in ogni riga, il valore assunto dalla funzione Y. Il valore assunto può essere indicato con T (per vero) e F (per falso) oppure con 1 (per vero) e 0 (per falso).

Noi faremo uso di questa seconda modalità.

A	B	C	Y
F	F	F	F
F	F	T	T
F	T	F	T
F	T	T	F
T	F	F	F
T	F	T	F
T	T	F	T
T	T	T	F

OPERATORI

L'algebra di Boole si basa su tre operatori: AND, OR e NOT (a cui corrispondono altrettante **porte logiche fondamentali**, usate negli schemi dei circuiti elettronici digitali).

AND

Prodotto logico (A and B)

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

In genere si usa l'operatore " · " che viene scritto espressamente, oppure omesso:

$$Y = A \cdot B \qquad \text{oppure} \qquad Y = AB$$

OPERATORI

OR

Somma logica (A or B)

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

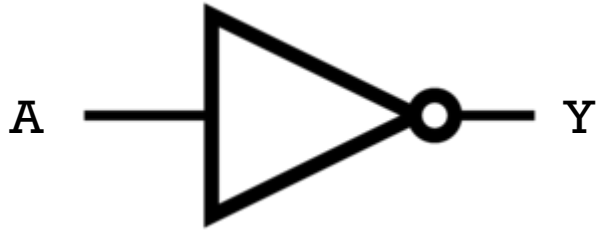
In genere si usa l'operatore " + " come nell'espressione:

$$Y = A + B$$

OPERATORI

NOT

Negazione logica (A negato)



A	Y
0	1
1	0

In genere si usa l'operatore " ¯ " (trattino in alto alla variabile cui fa riferimento) come nell'espressione:

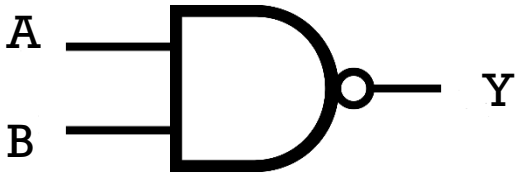
$$Y = \bar{A}$$

OPERATORI

Qualsiasi proposizione può essere scritta utilizzando la combinazione delle porte logiche fondamentali. Per semplificare gli schemi digitali o le proposizioni, però, spesso si fa uso delle seguenti **porte logiche derivate**.

NAND

Prodotto logico negato (A nand B)



A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

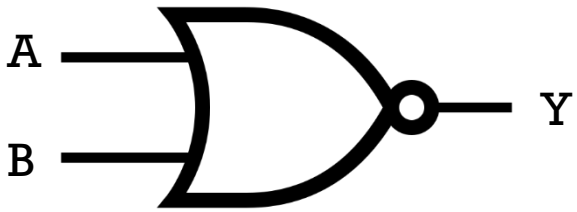
In genere si usa l'operatore "$\overline{\cdot}$" che viene scritto con il puntino, oppure senza, lasciando solo il trattino in alto:

$$Y = \overline{A \cdot B} \qquad \text{oppure} \qquad Y = \overline{AB}$$

OPERATORI

NOR

Somma logica negata (A nor B)



A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

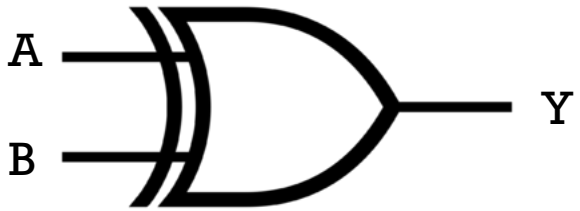
In genere si usa l'operatore "$\overline{\quad + \quad}$" come nell'espressione:

$$Y = \overline{A + B}$$

OPERATORI

XOR

OR esclusivo (A or esclusivo B o A diverso da B)



A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

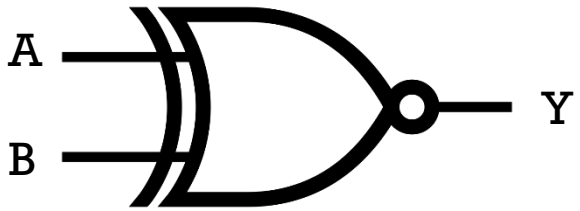
In genere si usa l'operatore $\oplus$ come nell'espressione:

$$Y = A \oplus B$$

OPERATORI

XNOR

NOR esclusivo (A nor esclusivo B)



A	B	Y
0	0	1
0	1	0
1	0	0
1	1	1

In genere si usa l'operatore $\overline{\oplus}$ come nell'espressione:

$$Y = \overline{A \oplus B}$$

FUNZIONI

Abbiamo visto che una funzione booleana (o funzione logica) è una legge F che presa una $n-upla$ di variabili booleane $(x_1, x_2, \dots, x_n)$ fa corrispondere ad essa il valore:

$$F(x_1, x_2, \dots, x_n) = \begin{cases} true \\ o \\ false \end{cases}$$

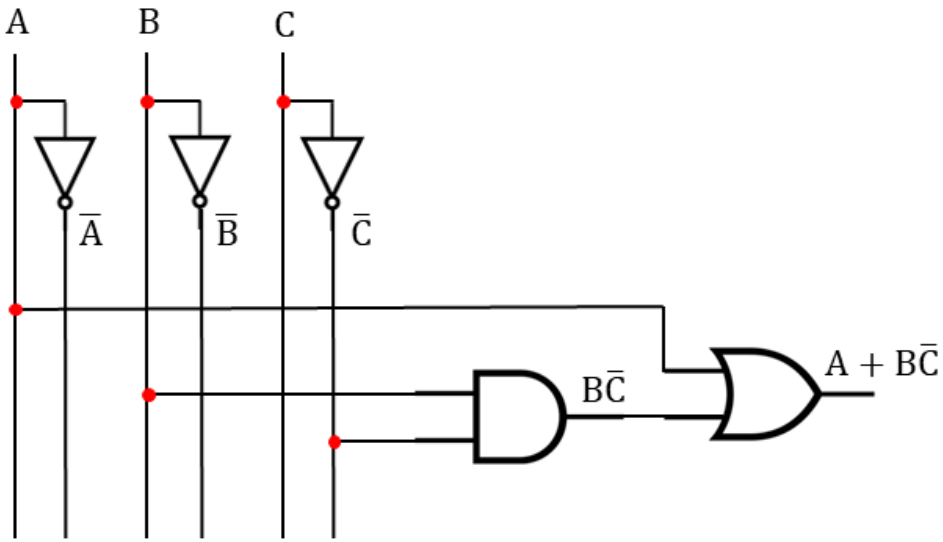
Inoltre, il suo comportamento può essere descritto attraverso una **tavola di verità** o attraverso un **circuito digitale**.

ESEMPIO

Consideriamo, ad esempio, la seguente funzione booleana:

$$F = A + B\bar{C}$$

La *tavola di verità* ed il *circuito digitale combinatorio* corrispondenti saranno:



A	B	C	F
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

MINTERMINI E MAXTERMINI

Data una tavola di verità e presa una riga, si definisce *mintermine* quello ottenuto moltiplicando tra loro le variabili booleane di input, presenti in tale riga, invariate, se assumono valore 1 o negate se assumono valore 0.

Si definisce, invece, *maxtermine* quello ottenuto sommando tra loro le variabili booleane di input, presenti sempre nella riga considerata, questa volta prese invariate, se assumono valore 0, o negate se assumono valore 1.

In una tavola di verità con n variabili, come detto, si possono avere 2^n input, pertanto si avranno altrettanti mintermini e maxtermini.

MINTERMINI E MAXTERMINI

Se consideriamo nuovamente la seguente tavola di verità:

A	B	C	F
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

in essa, presa la terza riga (quella con 010 in input), il mintermine corrispondente sarà:

$$\bar{A}B\bar{C}$$

mentre il maxtermine corrispondente sarà:

$$A + \bar{B} + C$$

FORME CANONICHE

Data una generica funzione booleana e considerata la sua rappresentazione attraverso una tavola di verità, essa può essere espressa nella *prima forma canonica* (o forma disgiuntiva):

- considerando solo le righe in cui essa assume in uscita il valore 1;
- costruendo i mintermini corrispondenti;
- sommando logicamente (OR) tutti i mintermini ottenuti al punto precedente.

FORME CANONICHE

Essa può essere, invece, espressa nella *seconda forma canonica* (o forma congiuntiva):

- considerando solo le righe in cui essa assume in uscita il valore 0;
- costruendo i maxtermini corrispondenti;
- moltiplicando logicamente (AND) tutti i maxtermini ottenuti al punto precedente.

ESEMPIO

Si consideri sempre la tavola di verità vista in precedenza:

A	B	C	F
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

I termini che portano al valore $F = 1$ sono:

010, 100, 101, 110 e 111

Considerando gli input in cui essa vale 1, costruendo i mintermini corrispondenti e sommando logicamente tra loro tali termini, si ottiene:

$$F = \bar{A}B\bar{C} + A\bar{B}\bar{C} + A\bar{B}C + AB\bar{C} + ABC$$

che è la prima forma canonica di F (si provi a costruire la seconda forma come esercizio).

FUNZIONI

Si noti, però, come la rappresentazione di F alla quale siamo appena giunti con questo procedimento, ovvero:

$$F = \bar{A}B\bar{C} + A\bar{B}\bar{C} + A\bar{B}C + AB\bar{C} + ABC$$

sia di gran lunga più complessa rispetto a quella dalla quale siamo partiti nella slide in cui abbiamo costruito anche il circuito digitale equivalente, ovvero:

$$F = A + B\bar{C}$$

Ciò è dovuto al fatto che, per ciascuna funzione booleana **esistono più rappresentazioni equivalenti.** Cerchiamo di capire, allora, in che modo è possibile determinare la forma ridotta, trasformando la prima nella seconda.

PROPRIETÀ DELL'ALGEBRA DI BOOLE

Nell'algebra booleana valgono diversi assiomi, utilizzando i quali è possibile manipolare le espressioni logiche al fine, ad esempio, si semplificarle:

$A + 0 = A$	proprietà di identità
$A + 1 = 1$	proprietà di assorbimento
$A + \bar{A} = 1$	proprietà dell'inverso
$A + A = A$	proprietà dell'idempotenza
$\bar{\bar{A}} = A$	proprietà della doppia negazione

PROPRIETÀ DELL'ALGEBRA DI BOOLE

Principio di dualità

Una proprietà è valida se e solo se è valida la sua duale, ovvero quella che si ottiene scambiando tra loro somma con prodotto e 1 con 0.

Applicando tale principio, le proprietà precedenti possono essere estese al prodotto logico:

$A \cdot 1 = A$	proprietà di identità
$A \cdot 0 = 0$	proprietà di assorbimento
$A \cdot \bar{A} = 0$	proprietà dell'inverso
$A \cdot A = A$	proprietà dell'idempotenza

PROPRIETÀ DELL'ALGEBRA DI BOOLE

Valgono, inoltre, le seguenti proprietà e teoremi:

Proprietà commutativa

$$A + B = B + A \qquad A \cdot B = B \cdot A$$

Proprietà associativa

$$A + (B + C) = (A + B) + C \qquad A \cdot (B \cdot C) = (A \cdot B) \cdot C$$

Proprietà distributiva

$$A \cdot (B + C) = (A \cdot B) + (A \cdot C) \qquad A + (B \cdot C) = (A + B) \cdot (A + C)$$

PROPRIETÀ DELL'ALGEBRA DI BOOLE

O ancora…

Teoremi di De Morgan

$$\overline{A + B} = \bar{A} \cdot \bar{B} \qquad \overline{A \cdot B} = \bar{A} + \bar{B}$$

I° teorema dell'assorbimento

$$A + A \cdot B = \mathrm{A} \qquad A \cdot (A + B) = \mathrm{A}$$

II° teorema dell'assorbimento

$$A + \bar{A} \cdot B = \mathrm{A} + \mathrm{B} \qquad A \cdot (\bar{A} + B) = \mathrm{A} \cdot B$$

ESEMPIO

Utilizzando le proprietà appena illustrate, vediamo in che modo è possibile semplificare la precedente funzione:

$$F = \bar{A}B\bar{C} + A\bar{B}\bar{C} + A\bar{B}C + AB\bar{C} + ABC$$

Modificando l'ordine e applicando la proprietà dell'idempotenza $A + A = A$ al penultimo termine possiamo scrivere:

$$F = A\bar{B}\bar{C} + A\bar{B}C + ABC + AB\bar{C} + AB\bar{C} + \bar{A}B\bar{C}$$

Tra i primi due termini possiamo mettere in evidenza $A\bar{B}$, tra terzo e quarto AB e tra quinti e sesto $B\bar{C}$, ottenendo:

$$F = A\bar{B} + AB + B\bar{C} = A(B + \bar{B}) + B\bar{C} = A + B\bar{C}$$

avendo messo in evidenza A tra i primi due termini e semplificato.

MINIMIZZAZIONE DI FUNZIONI BOOLEANE

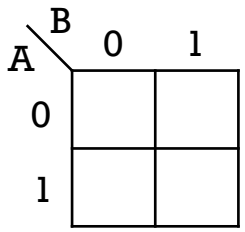
Riscrivere una funzione booleana nella forma minima equivalente, significa semplificare (e quindi ridurne il costo di realizzazione) il circuito digitale con il quale implementarla, ovvero ridurre il numero di componenti elettronici da utilizzare, ma anche il livello (ovvero il numero massimo di componenti da attraversare tra input e output, fatta eccezione per le porte NOT).

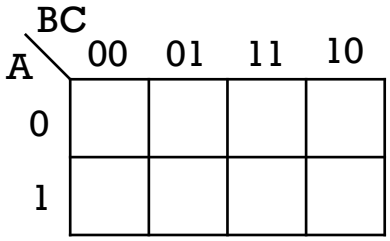
Procedere alla semplificazione ricorrendo alle proprietà nel modo visto nell'ultimo esempio, però, non è sempre una operazione semplice da compiere.

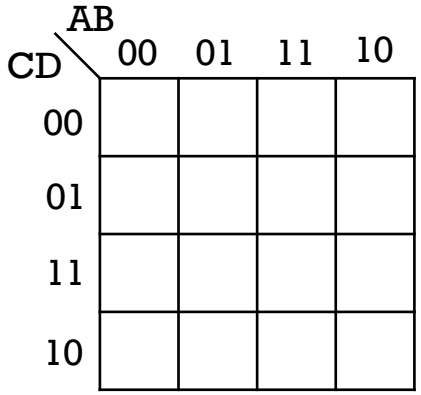
Vediamo adesso un altro metodo che fa uso di una differente forma di rappresentazione di una funzione booleana, ovvero delle *Mappe di Karnaugh*.

MINIMIZZAZIONE DI FUNZIONI BOOLEANE

Una Mappa di Karnaugh può essere considerata come una variante delle tavole di verità. Vediamo come costruirle a 2, a 3 o a 4 variabili booleane:







È importante notare che le mappe sono costruire in modo tale che, nello spostamento tra righe o colonne adiacenti, solo una variabile per volta cambia il proprio valore.

MINIMIZZAZIONE DI FUNZIONI BOOLEANE

Consideriamo ancora una volta la seguente tavola di verità:

La mappa di karnaugh associata ad una funzione, si costruisce riportando il valore 1 in corrispondenza dei mintermini che assumono tale valore (gli 0 si possono anche omettere):

A	B	C	F
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

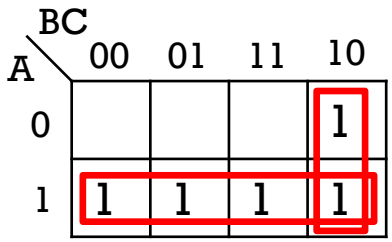
A \ BC	00	01	11	10
0				1
1	1	1	1	1

MINIMIZZAZIONE DI FUNZIONI BOOLEANE

Disegnata la mappa di Karnaugh, per semplificare occorre:

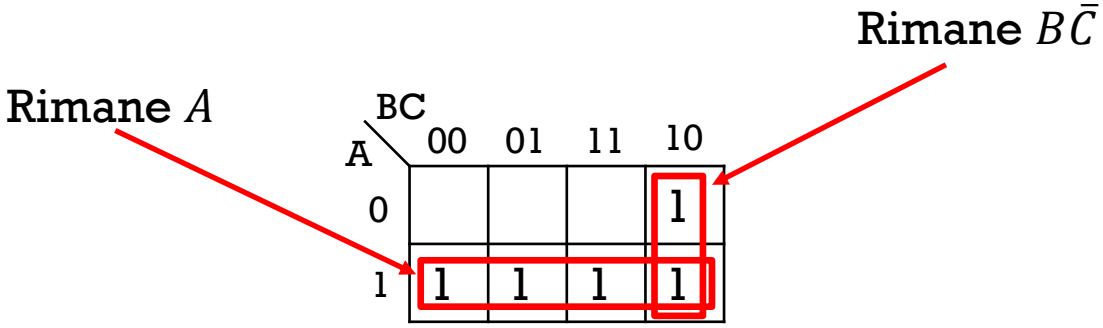
- raggruppare i valori 1 adiacenti (considerando sempre un numero di valori che sia potenza di 2, ovvero 2, 4, 8,…);
- considerare solo i gruppi di bit essenziali, ovvero la combinazione minima tra quelli che contengono valori di uscita che non siano già coperti da altri gruppi.

Nel nostro esempio avremo quindi:



MINIMIZZAZIONE DI FUNZIONI BOOLEANE

Fatto questo, per ciascun gruppo si mantiene solo la variabile (o le variabili) booleane il cui valore non cambia per tutti i mintermini del gruppo. Ciascuna variabile verrà presa negata se ha valore 0, o invariata se ha valore 1.

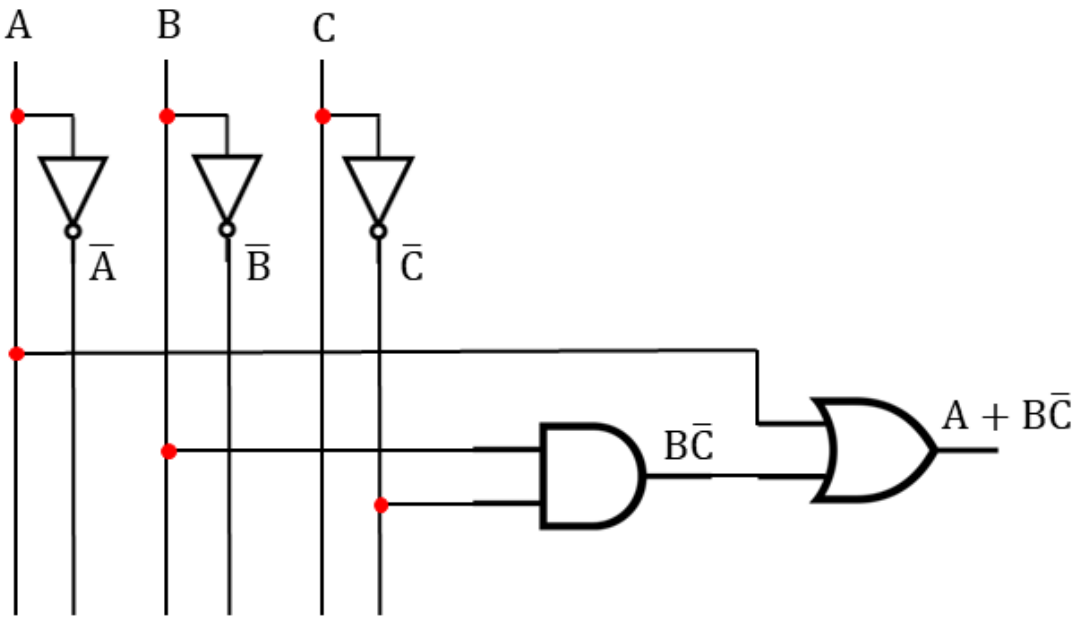


Questi nuovi termini semplificati, andranno sommati logicamente tra loro:

$$F = A + B\bar{C}$$

CIRCUITO DIGITALE

Ottenuta la funzione booleana equivalente a quella di partenza, ma in forma minima, si può procedere con il disegno dello schema relativo al circuito digitale combinatorio che la implementa:



Un circuito digitale è detto *combinatorio* quando è privo di retroazioni (il segnale viaggia a senso unico dall'input verso l'output), mentre è detto *sequenziale* quando presenta retroazioni (l'uscita di una porta può puntare a porte precedenti).

ARCHITETTURA DEI CALCOLATORI

LA MACCHINA DI VON NEUMANN

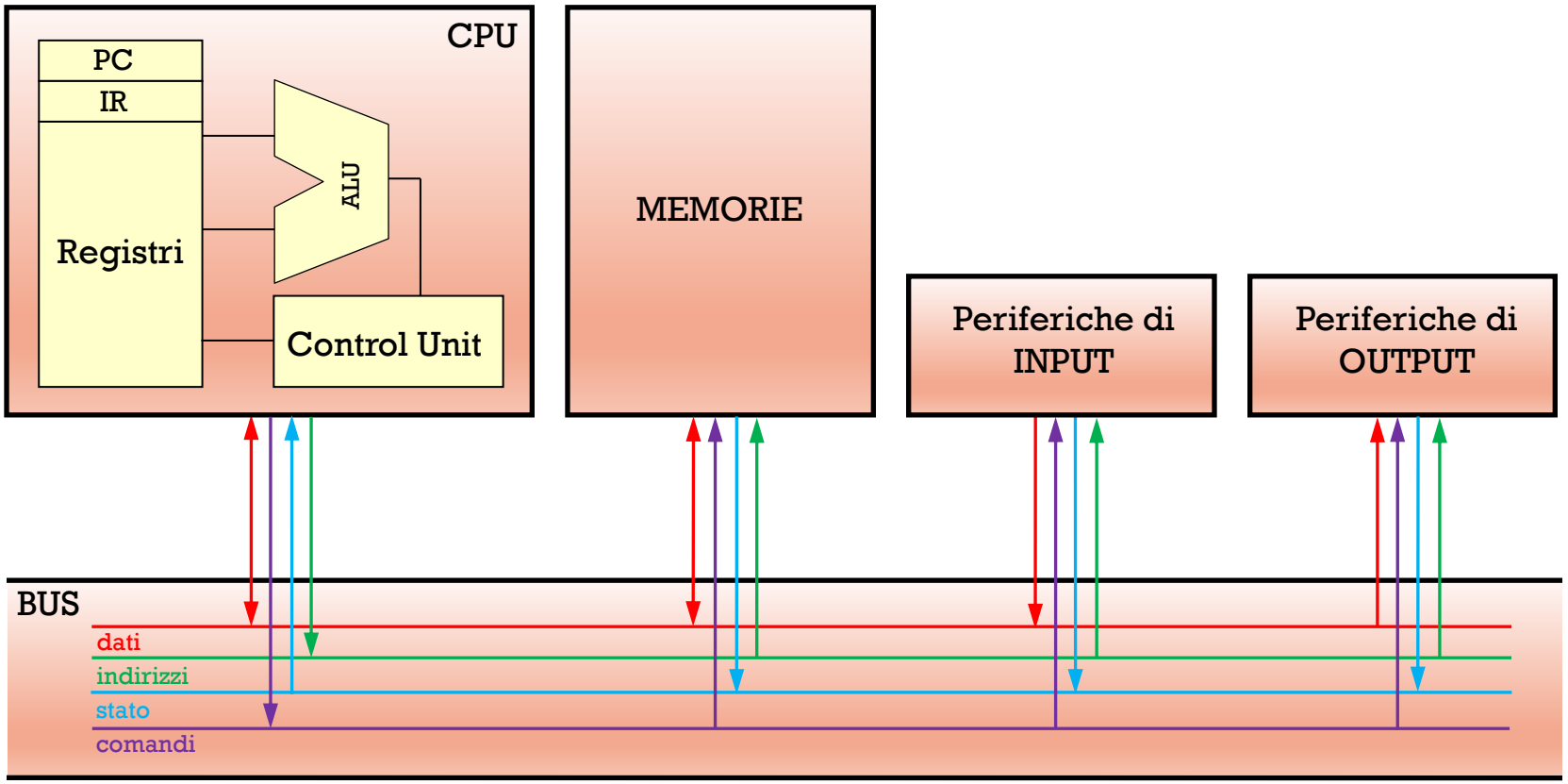
In un computer possiamo distinguere tre tipi di unità funzionali:

- Processore (fornisce la capacità di elaborazione delle informazioni);
- Memoria (centrale e di massa);
- Dispositivi di input/output.

Le parti comunicano attraverso un canale detto BUS, costituito da un insieme di linee elettriche digitali.

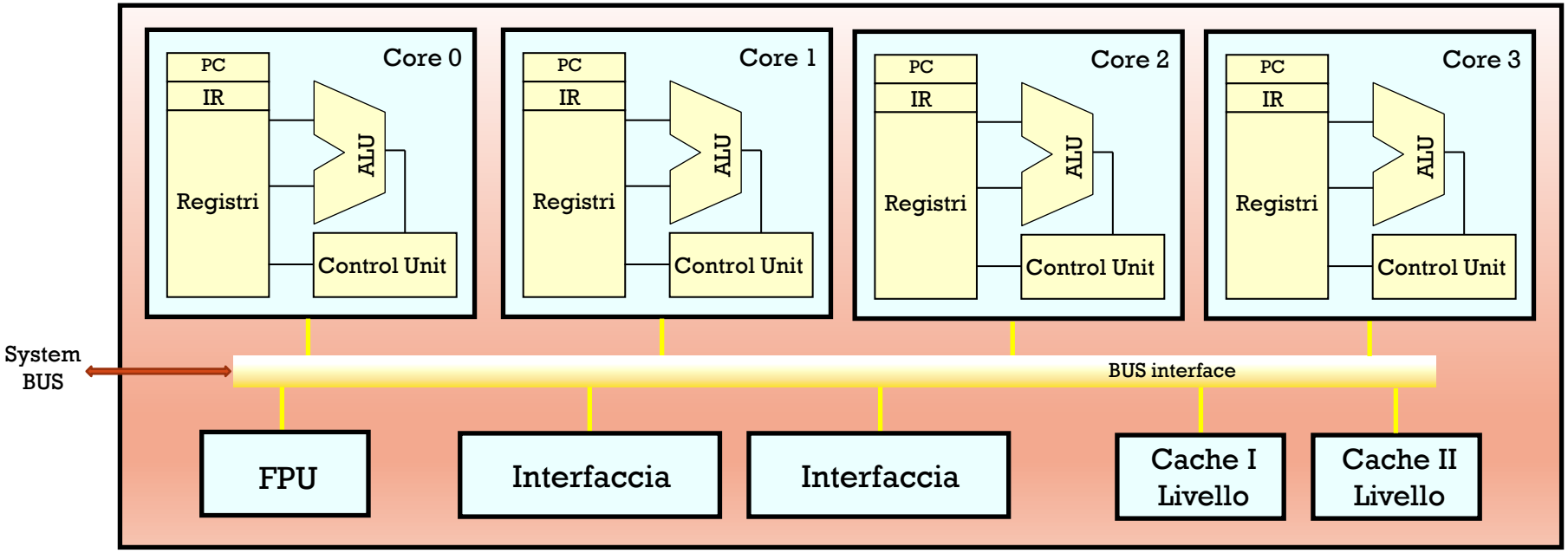
La maggior parte degli elaboratori moderni (compresi gli smartphone ed i tablet) sono basati sulla macchina di Von Neumann.

SCHEMA A BLOCCHI DI UN ELABORATORE



PROCESSORE

Oggi un processore non viene più identificato con una singola CPU, ma con più CPU (processori multicore). Sebbene sia più complesso, uno schema può essere approssimato come segue:



PROCESSORE

Dunque, è composto da blocchi con funzionalità diverse:

- 2^n CPU (Central Processing Unit);
- FPU (Floating Point Unit);
- Cache L1 (o di primo livello);
- Interfacce varie.

Se tutto questo è integrato all'interno di un unico chip, prende anche il nome di microprocessore. I moderni processori sono talvolta dotati anche di una Cache L2 (o di secondo livello).

Di solito le cache L1 sono più piccole (128KB o 256 KB) quindi integrate con lo stesso chip, mentre le cache L2 sono più grandi (valori tipici vanno da 512KB fino a 8MB o anche più), pertanto talvolta sono implementate in un circuito esterno (quindi con velocità di accesso minore).

In questo modo si riescono a gestire esigenze differenti in fase di elaborazione dei programmi.

CENTRAL PROCESSING UNIT (CPU)

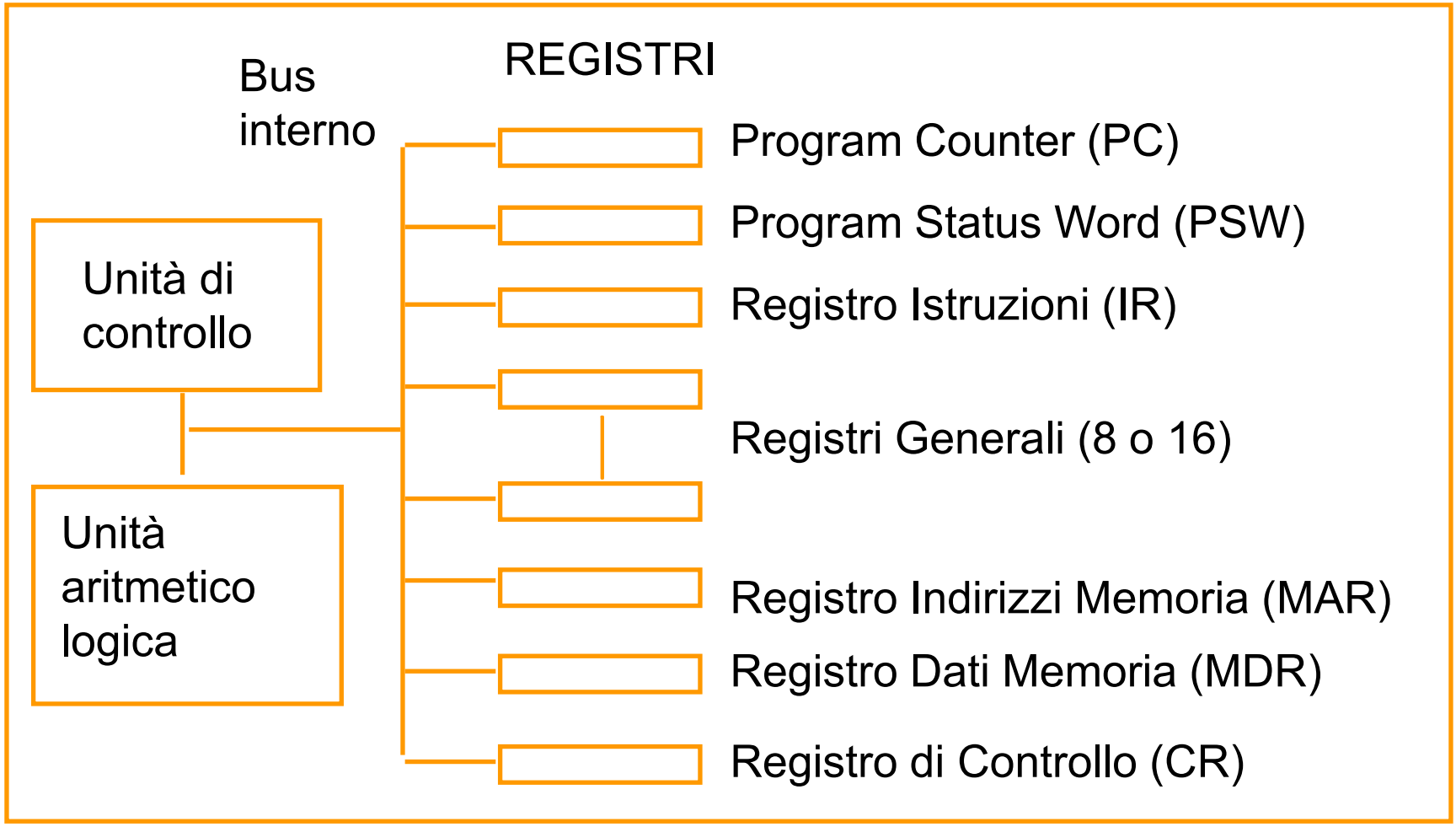
Svolge tutte le operazioni di:

- elaborazione numerica;
- controllo e coordinamento di tutte le attività.

Ciascuna CPU si suddivide in:

- Unità logico-aritmetica (ALU);
- Unità di controllo (CU);
- Registri.

COMPONENTI DELLA CPU



FLOATING POINT UNIT (FPU)

L'unità di calcolo in virgola mobile è una parte del processore che si contraddistingue per essere specializzata nell'esecuzione di calcoli matematici in virgola mobile (dotati quindi di parte intera e parte decimale).

La maggior parte delle operazioni di calcolo svolte dalla CPU sono operazioni aritmetiche di base (somme, sottrazioni, moltiplicazioni e divisioni) effettuate sull'insieme dei numeri naturali o sull'insieme dei numeri relativi, utilizzando la ALU.

Quando invece è necessario effettuare calcoli più complessi (esponenziali, trigonometrici, etc.), lavorando con numeri reali, entra in gioco la FPU.

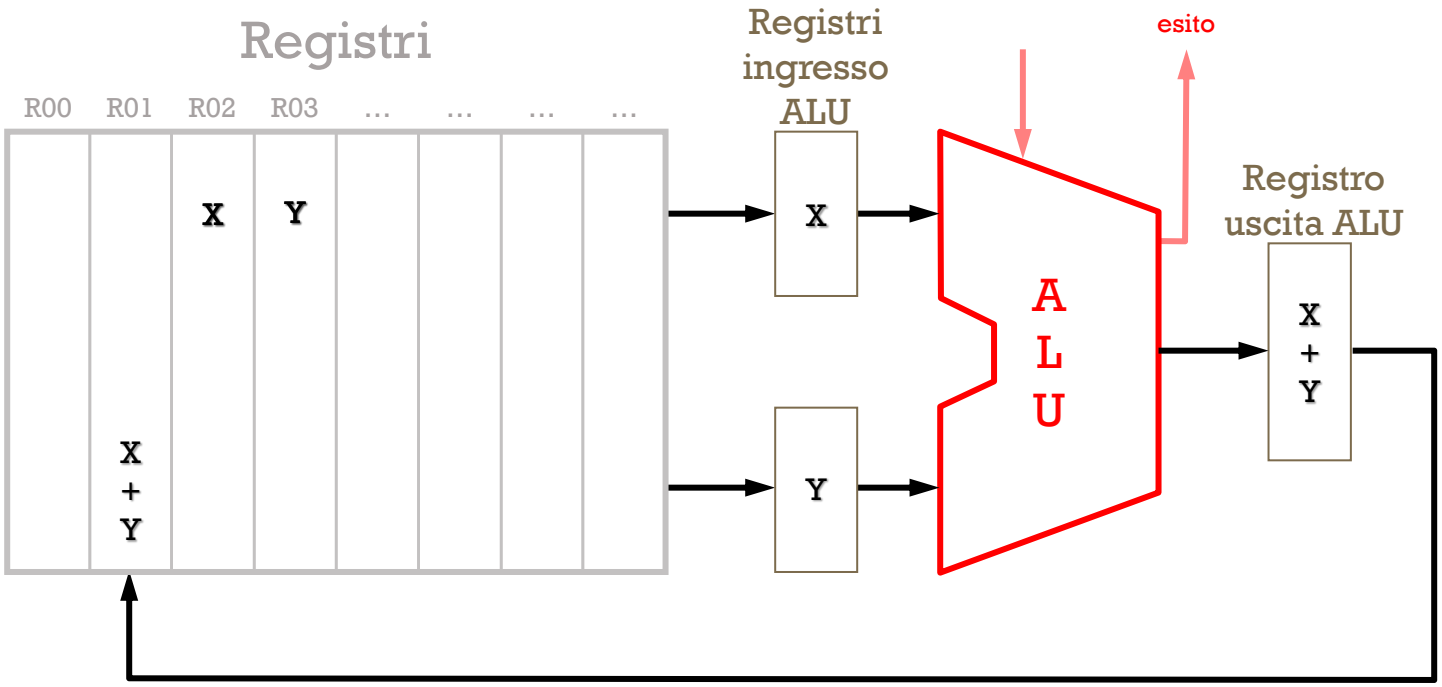
ARITHMETIC LOGIC UNIT (ALU)

Svolge tutti i calcoli logici ed aritmetici operando direttamente con i registri.

In realtà, è costituita da circuiti elettronici in grado di eseguire solamente la somma di due numeri binari contenuti in due registri, oppure il confronto tra due numeri, sempre presenti in altrettanti registri.

Le altre operazioni vengono eseguite come combinazione delle sole operazioni indicate sopra (così come visto in precedenza).

ARITHMETIC LOGIC UNIT (ALU)



ARITHMETIC LOGIC UNIT (ALU)

Per quanto riguarda le operazioni logiche una ALU, di solito, è un grado di implementare le porte AND, OR, NOT e XOR.

Von Neumann, nel suo modello, ritenne ragionevole introdurre una unità specializzata allo scopo di compiere le operazioni elencate.

Si noti come la stessa ALU venga costruita facendo uso proprio delle porte logiche fondamentali AND, OR e NOT.

REGISTRI

Sono delle piccole porzioni di memoria locale utilizzata per il salvataggio di:

- dati acquisiti dalla memoria centrale o dalle unità di input;
- risultati delle operazioni eseguite dall' ALU.

I registri sono in numero limitato (tipicamente non oltre 256).

Sono unità di memoria estremamente veloci, di dimensione pari a 16, 32 o 64 bit (tale valore dipende dall'architettura del processore).

REGISTRI

I registri contengono dati ed istruzioni che vengono immediatamente elaborati.

Esistono due tipi di registri:

- registri speciali utilizzati dalla CU per scopi particolari;
- registri di uso generale (solitamente per operazioni aritmetiche).

IL SEGNALE DI CLOCK

Ogni elaboratore contiene un circuito di temporizzazione (detto clock), che genera un riferimento temporale comune per tutti gli elementi del sistema.

- T = **periodo** di clock
- f = **frequenza** di clock (= 1/T)

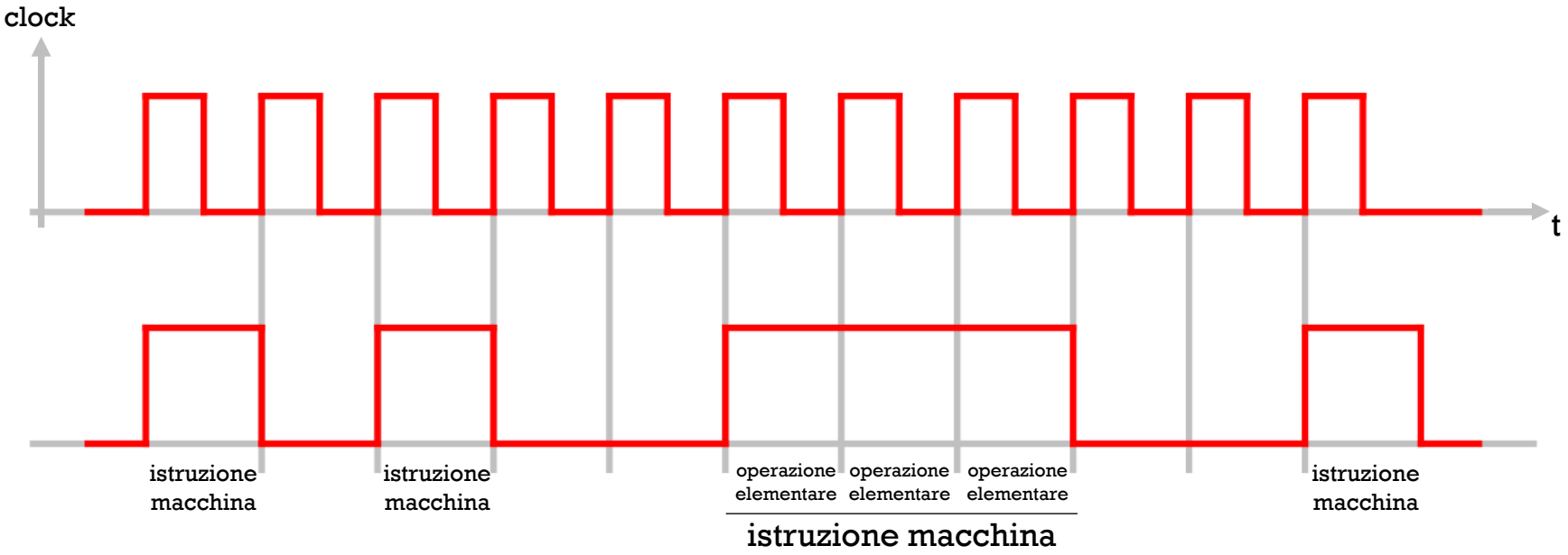
La frequenza di clock è misurata in Hertz (Hz).

TEMPISTICA DELLE ISTRUZIONI

Un **ciclo-macchina** è il tempo necessario per svolgere una *operazione elementare*. È un multiplo del periodo del clock.

Un'*istruzione macchina* è ottenuta da una sequenza di operazioni elementari. Dunque, l'esecuzione di un'istruzione macchina richiede un numero intero di cicli macchina, variabile a seconda del tipo di istruzione.

TEMPISTICA DELLE ISTRUZIONI



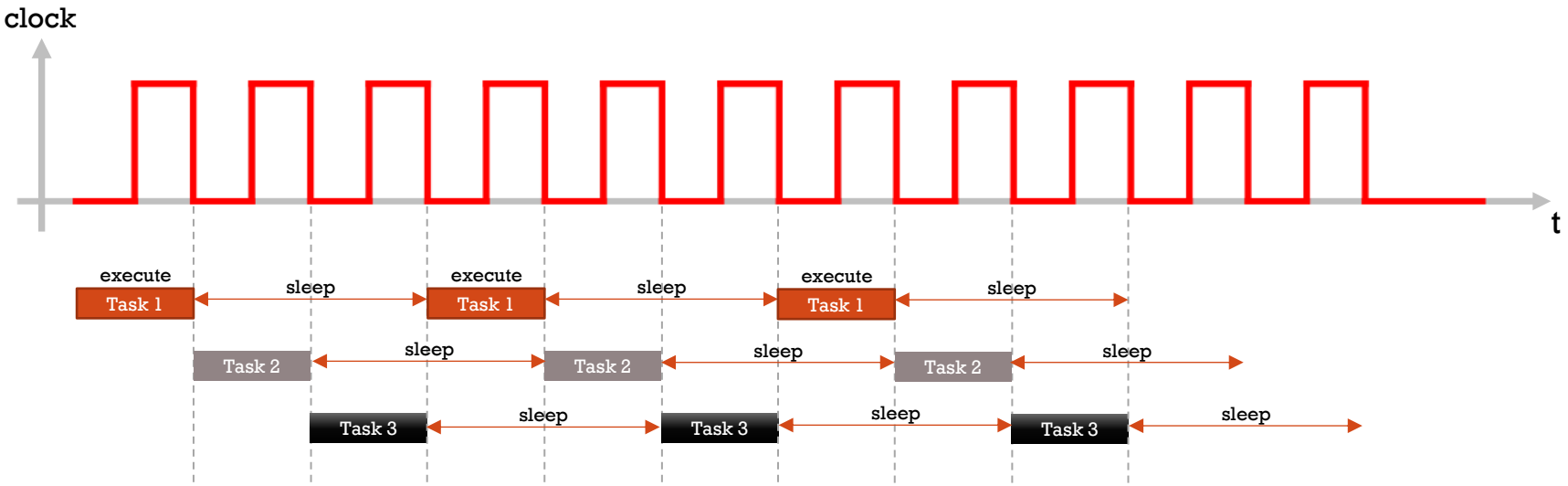
Nell'esempio sopra, le prime due istruzioni macchina e l'ultima richiedono esattamente una operazione elementare ciascuna, la terza istruzione, invece, richiede l'esecuzione di 3 operazioni elementari. Si sta ipotizzando che 1 ciclo-macchina coincida con il periodo del clock.

VELOCITÀ DEL MICROPROCESSORE

La velocità di elaborazione di un processore dipende dalla frequenza del clock.

I processori attuali hanno valori di frequenza di clock dell'ordine dei GHz (Giga Hertz), il che significa che un processore con tali caratteristiche è in grado di eseguire miliardi di operazioni elementari al secondo.

TIME SHARING



Nell'esempio sopra abbiamo ipotizzato che la fase di esecuzione di ciascuno dei 3 task preveda il coinvolgimento di una singola *operazione elementare*

MEMORIA

Viene utilizzata per conservare dati e programmi. Si suddivide in:

- memoria principale;

  Memoria in grado di conservare dinamicamente dati e programmi che il processore sta utilizzando.

- memoria di massa.

  Memoria in grado di conservare permanentemente dati e programmi degli utenti.

MEMORIA PRINCIPALE

La memoria *principale* (o centrale) è utilizzata per ospitare, in modo *temporaneo,* i dati ed i programmi che gli utenti richiedono di eseguire e/o elaborare.

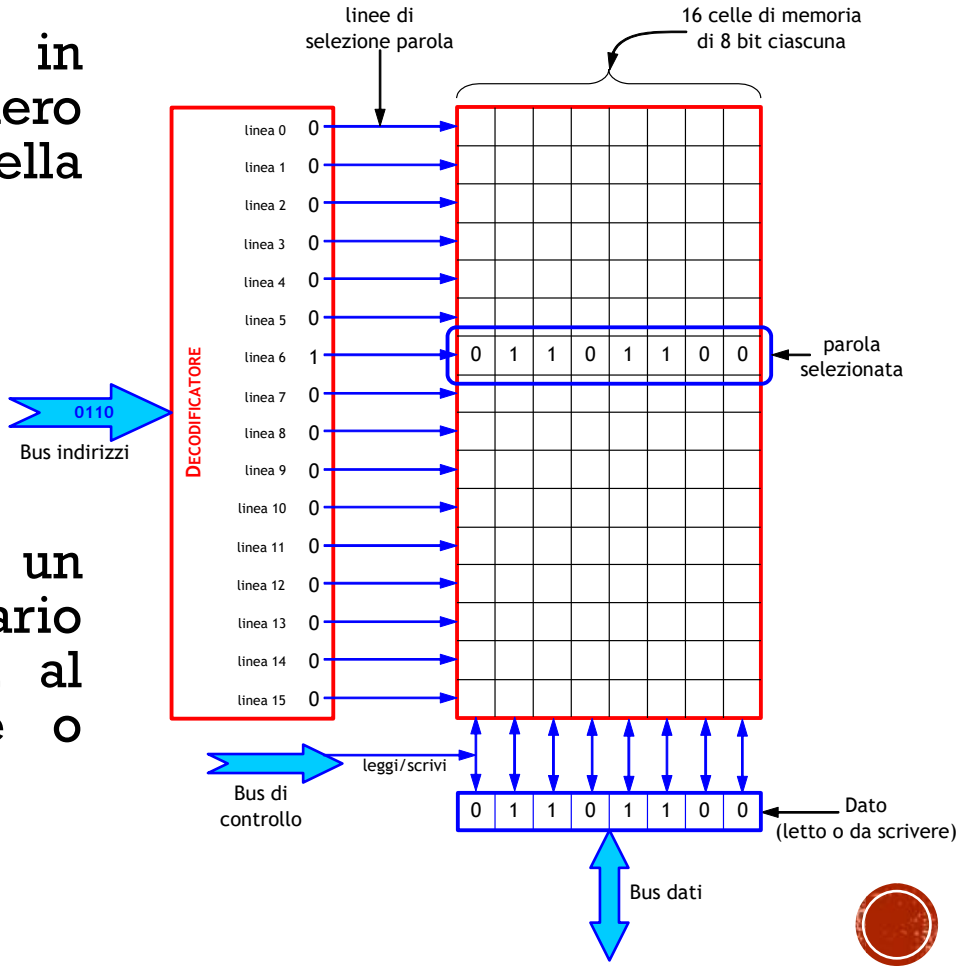
Prevedono *piccole capacità* di memorizzazione e sono in genere costituite da RAM, memorie cache o registri. Alcune memorie centrali, invece, contengono in modo permanente informazioni necessarie al funzionamento dell'hardware (ROM).

Quando si vuole eseguire un programma, esso viene sempre caricato nella memoria principale (insieme agli eventuali dati da elaborare).

ORGANIZZAZIONE DELLA MEMORIA PRINCIPALE

La memoria è organizzata in *celle*, ciascuna grande un numero di bit pari all'architettura della CPU, tra loro indipendenti.

Ad ogni cella è associato un *indirizzo,* cioè un numero binario progressivo a partire da 0, al quale puntare per leggere o scrivere dati.

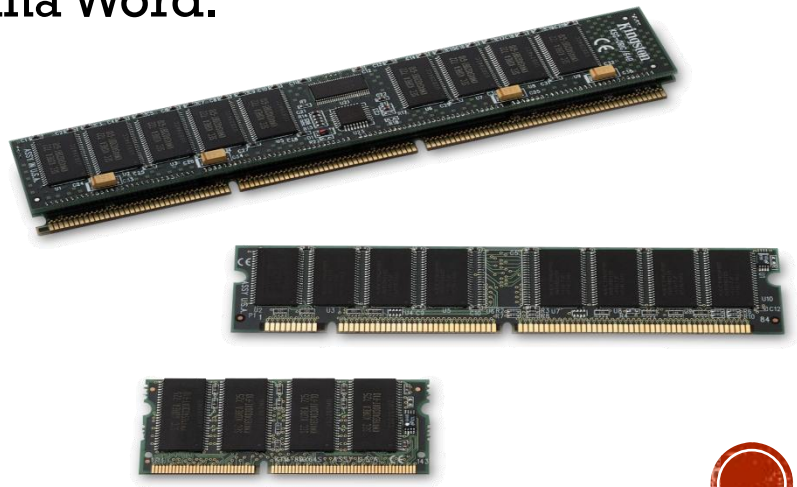


RANDOM ACCESS MEMORY (RAM)

È una memoria *volatile* basata su circuiti digitali, a disposizione della CPU per memorizzare programmi e dati necessari allo svolgimento delle elaborazioni richieste da un utente. A fine elaborazione, lo spazio occupato viene rilasciato.

Il tempo di accesso a tale memoria è il tempo necessario per leggere o scrivere l'informazione in una Word.

Il concetto «accesso casuale» indica che ogni cella ha lo stesso tempo di accesso, indipendentemente dall'indirizzo ad essa associato.



CARATTERISTICHE DELLA RAM

La dimensione di una RAM varia a seconda del tipo di computer e viene espressa in GB.

Le dimensioni tipiche della RAM di un calcolatore moderno vanno da 1GB a 16GB. La RAM di un Server, invece, può arrivare fino a 512GB.

Il tempo di accesso è dell'ordine di poche decine di nano-secondi (10^{-9} sec).

La RAM è una memoria costosa (il costo va inteso in valuta/bit).

READ ONLY MEMORY (ROM)

È una memoria di *sola lettura* che viene scritta direttamente dal produttore del computer su circuiti appositi.

Viene utilizzata per contenere le informazioni di inizializzazione usate ogni volta che si accende l'elaboratore. In alcuni casi, contiene un software particolare (chiamato firmware), installato all'interno sempre dal produttore.

Esistono diversi tipi di ROM, nate come evoluzione di quella sopra:

- PROM (Programmable ROM);
- EPROM (Erasable Programmable ROM);
- EEPROM (Electrically Erasable Programmable ROM);
- EAROM (Electrically Alterable ROM);
- Flash ROM.

READ ONLY MEMORY (ROM)

Una *PROM* (evoluzione di una ROM) viene prodotta vuota e può essere scritta da un utente, in seguito, utilizzando un'apposita attrezzatura, una volta sola.

Una *EPROM* (evoluzione di una PROM), può essere cancellata per intero, tramite esposizione ai raggi UV e riscritta.

Una *EEPROM* (evoluzione di una EPROM), può essere cancellata in blocchi di bit (al fine di essere riscritta), tramite l'applicazione di una tensione di valore superiore a quella di lavoro.

Una *EAROM* (che è un tipo particolare di EEPROM) può essere modificata in un singolo bit, invece che in blocchi di bit).

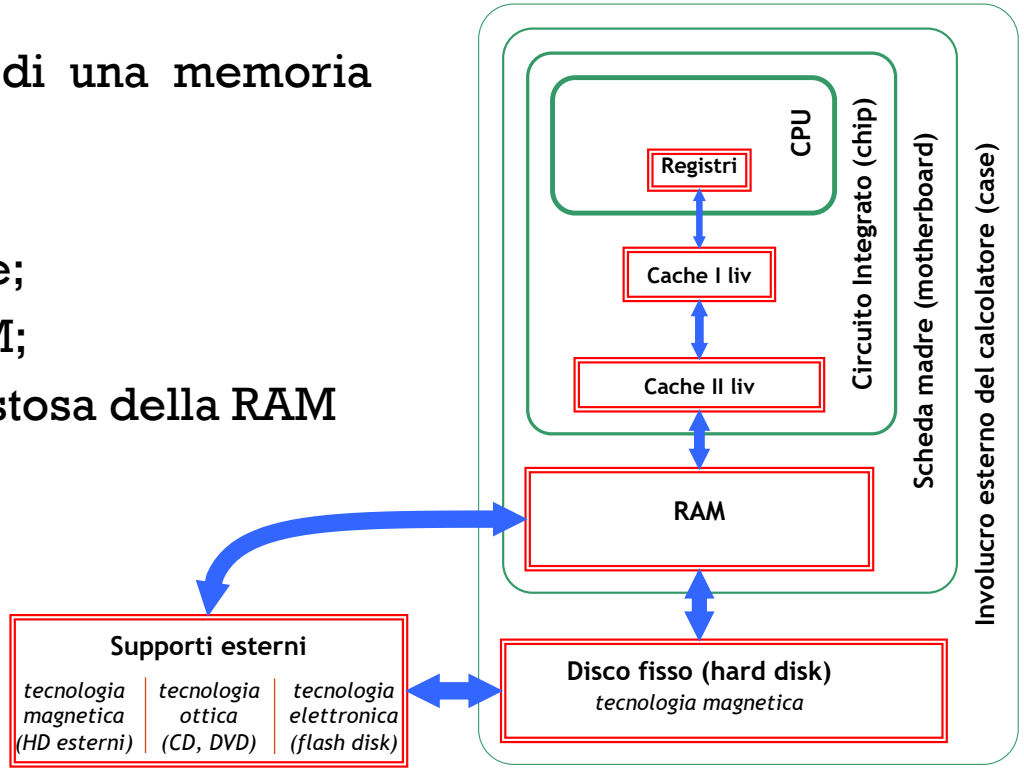
Una *Flash ROM*, infine, è una normale Flash utilizzata come ROM.

MEMORIA CACHE

Per migliorare le prestazioni di un computer si inserisce una memoria intermedia tra CPU e RAM, detta Cache.

Caratteristiche principali di una memoria Cache sono:

- è interna al processore;
- è più veloce della RAM;
- è di gran lunga più costosa della RAM ma anche più piccola.



MEMORIA CACHE

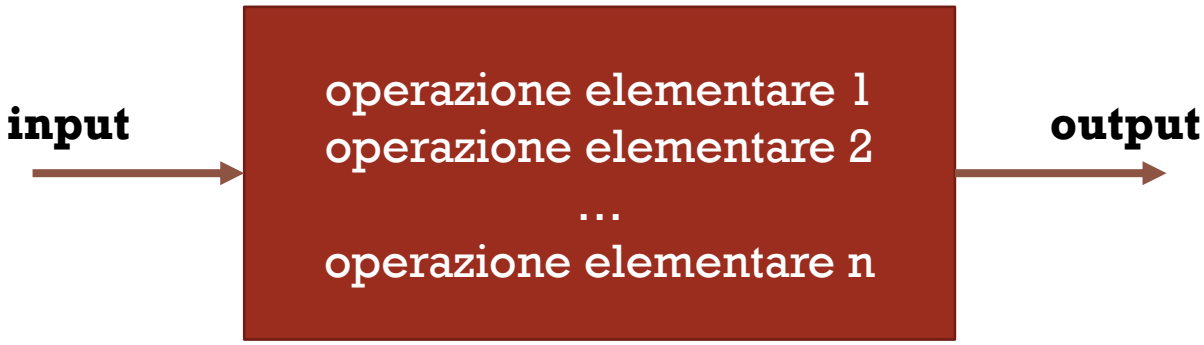
I dati e le istruzioni utilizzati con maggior frequenza vengono memorizzati nella *cache*, in modo da diminuirne il tempo di accesso ed aumentare, quindi, le prestazioni medie della CPU.

Diventa cruciale, dunque, il metodo per selezionare i dati e le istruzioni da inserire nella cache.

Le cache di I livello (L1), nei computer moderni, valgono 128KB o 256KB, mentre quelle di II livello (L2), hanno valori che vanno da 512 KB a 8÷ 12 MB.

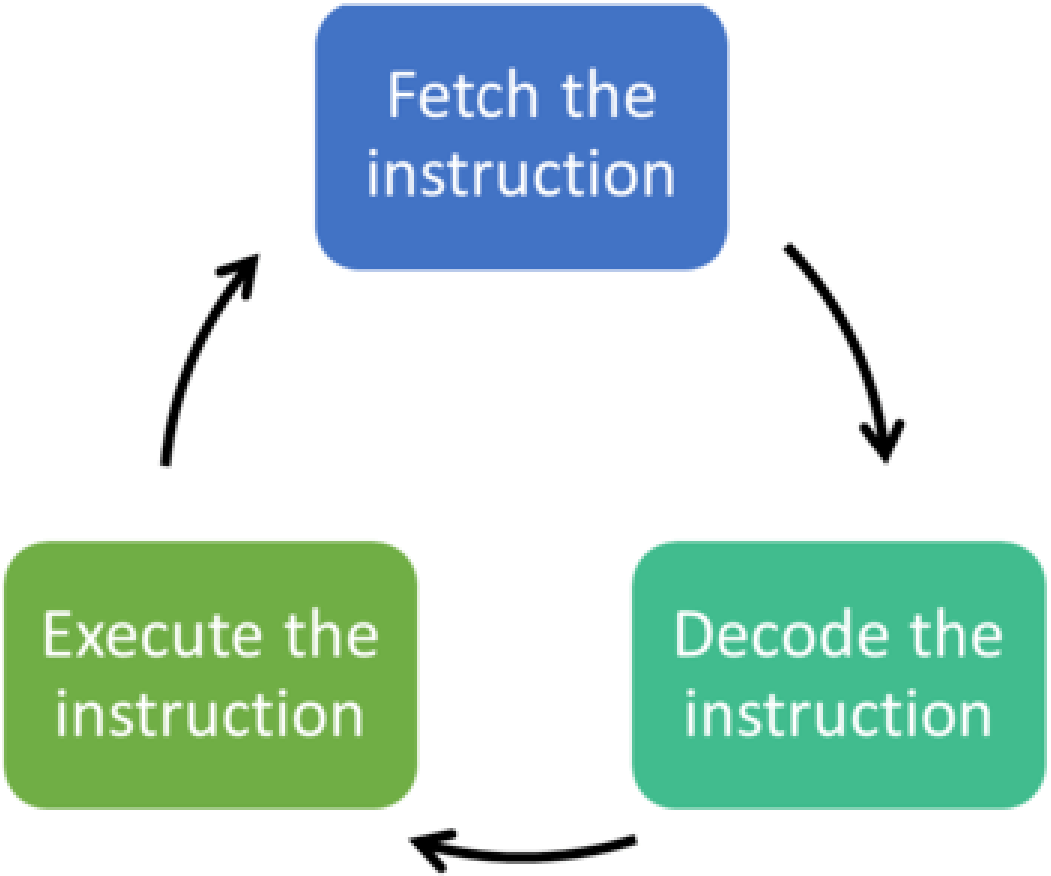
ALGORITMO

Si definisce *algoritmo* una sequenza ordinata e finita di operazioni elementari (istruzioni) che, presi un numero finito di dati in ingresso, dopo un tempo finito di elaborazione, produca un risultato univoco.



È importante osservare che le operazioni non devono poter essere scomposte in operazioni più semplici e non devono essere ambigue.

CICLO FETCH-DECODE-EXECUTE



CICLO FETCH-DECODE-EXECUTE

1. Programma e dati sono caricati in memoria centrale (RAM).

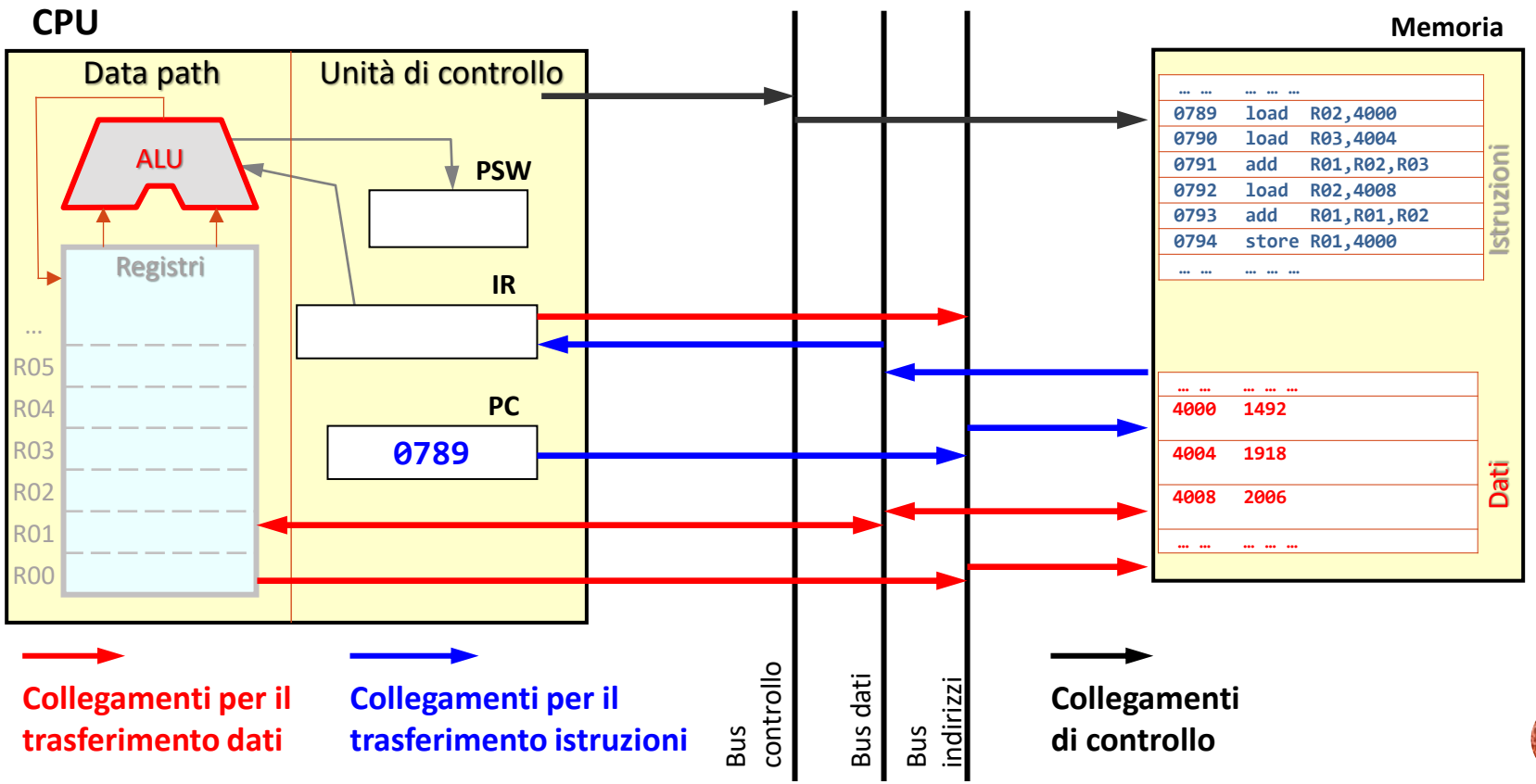
2. La CU ESTRAE la prima istruzione del programma dalla memoria (puntandola attraverso l'indirizzo, indicato nel registro PC) e la trasferisce nel registro IR, quindi incrementa di uno l'indirizzo in PC, predisponendosi alla successiva istruzione.

3. L'istruzione viene DECODIFICATA ed ESEGUITA, prelevando eventuali dati associati, dalla memoria, sempre attraverso l'indirizzo corrispondente. I dati vengono in genere trasferiti nei registri generali.

4. La CU passa all'istruzione successiva, attraverso l'indirizzo presente nel registro PC.

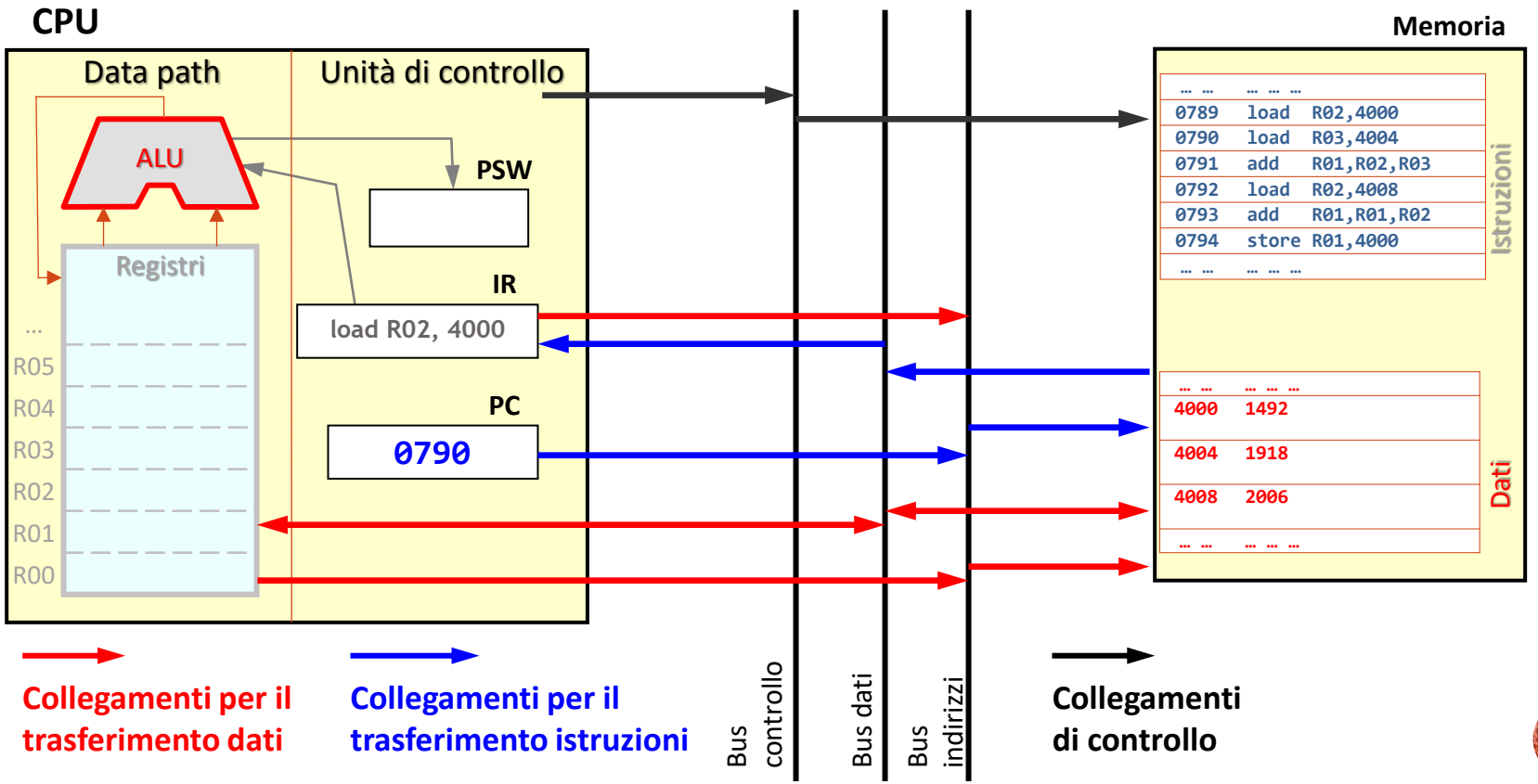
ESEMPIO

Di seguito, lo scenario dopo aver caricato programma e dati:



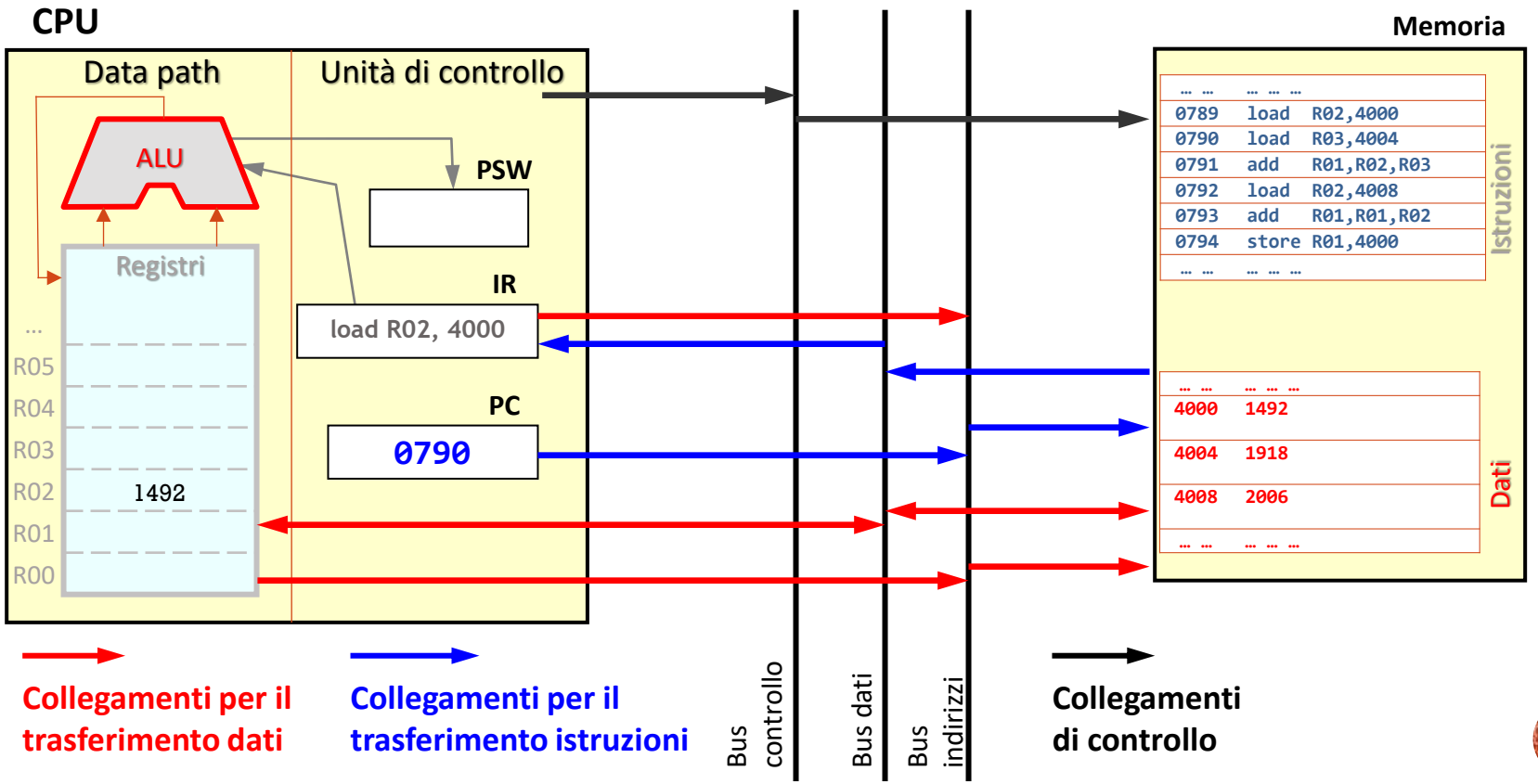
ESEMPIO

Vediamo un singolo ciclo: la CU ESTRAE l'istruzione all'indirizzo 0789, la trasferisce in IR, quindi incrementa PC:



ESEMPIO

La CU DECODIFICA l'istruzione in IR, (ovvero carica il dato all'indirizzo 4000 della RAM, nel registro R02) e la ESEGUE:



ESEMPIO

La CU prosegue con lo stesso lavoro visto nelle slides precedenti per il primo ciclo, fino a quando non completa l'esecuzione dell'intero blocco di istruzioni (l'algoritmo da eseguire), ovvero fino all'istruzione all'indirizzo 0794 della RAM.

Ad ogni ciclo, dunque, accede alla locazione di memoria associata all'indirizzo, estrae l'istruzione che vi trova dentro portandola nel registro IR (cancellando quindi l'eventuale precedente istruzione presente), incrementa l'indirizzo contenuto nel registro PC di una unità, decodifica l'istruzione in IR (eventualmente estraendo dati dalla RAM e portandoli in uno dei registri generali) e la esegue.

MEMORIA SECONDARIA

La memoria *secondaria* (o di massa) è utilizzata per conservare in modo *permanente* i dati ed i programmi degli utenti.

Richiede *grande capacità* di memorizzazione ed è principalmente costituita da dischi magnetici, CD-ROM, DVD, Flash Memory o nastri.

Quando si vuole eseguire un programma (così come visto nell'esempio precedente), esso viene sempre caricato dalla memoria di massa (insieme agli eventuali dati da elaborare) alla memoria principale.

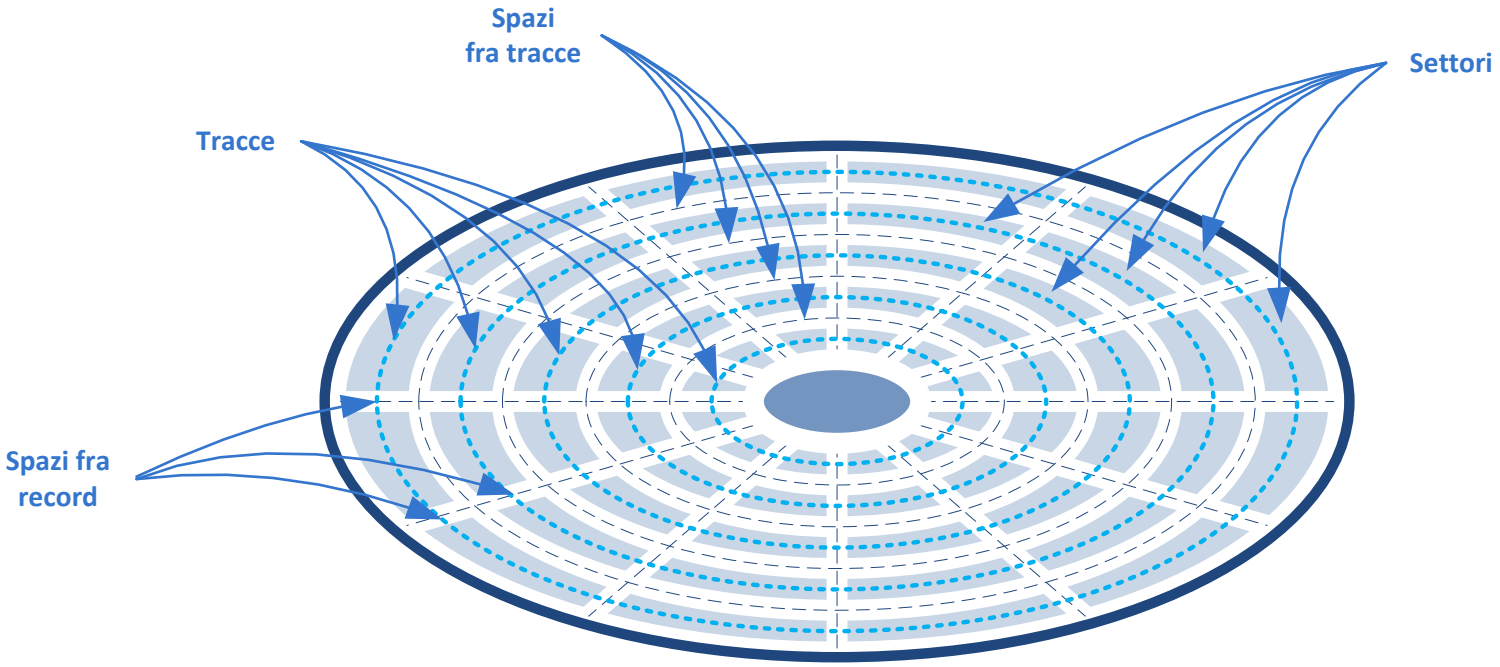
MEMORIA SECONDARIA

Tre tecnologie possibili:

- **Magnetica**
  - ✓ Dischi magnetici (Floppy-disk, Hard-disk)
    - Accesso casuale
    - Operazioni di Lettura/Scrittura
  - ✓ Nastri magnetici
    - Accesso sequenziale (legato alla posizione del dato)
    - Operazioni di Lettura/Scrittura
- **Ottica**
  - ✓ CD-ROM, DVD
    - Accesso casuale (tipicamente solo scrittura)
- **Flash memory**
  - ✓ Pen drive, schede di memoria (Sd, micro Sd, memory stick, compact flash, etc.
    - Accesso casuale (lettura/scrittura)

ORGANIZZAZIONE FISICA DEI DATI NEI DISCHI

Gli hard disk sono suddivisi in *tracce* concentriche e *settori*:



ORGANIZZAZIONE FISICA DEI DATI NEI DISCHI

Ogni *settore* è una "fetta" di disco. I settori suddividono ogni *traccia* in porzioni di circonferenza dette *blocchi*.

La suddivisione di un disco in tracce e settori viene indicata con il termine *formattazione*.

Per effettuare un'operazione di lettura o di scrittura su un blocco, è necessario che la testina raggiunga il blocco desiderato.

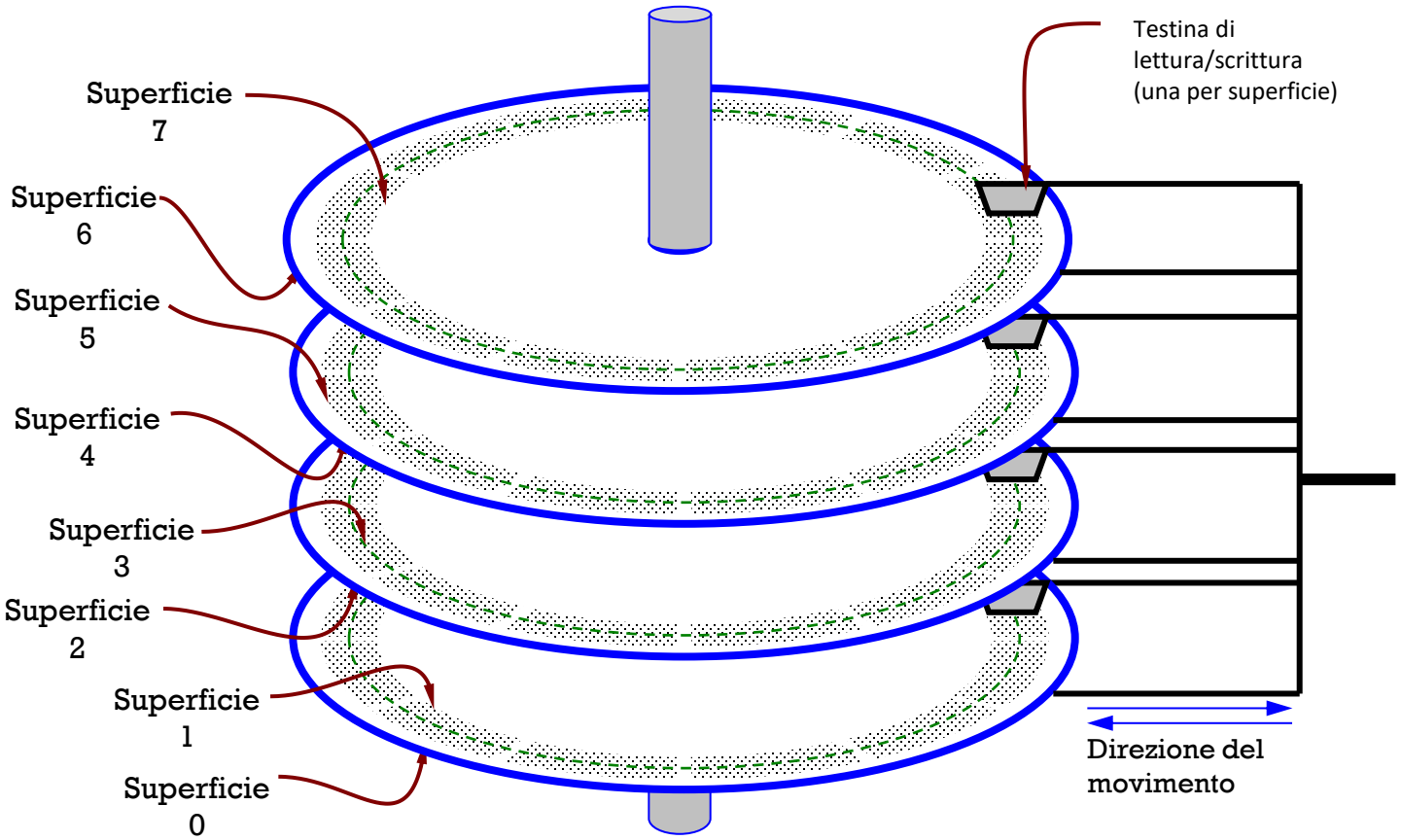
Ciascun disco può essere gestito in due modalità:

- Single-side.
- Double-side.

Per aumentare la capienza, spesso si montano più dischi, double-side, all'interno dello stesso case.

ORGANIZZAZIONE FISICA DEI DATI NEI DISCHI

Le tracce in grigio formano un "cilindro"



ORGANIZZAZIONE FISICA DEI DATI NEI DISCHI

Il *tempo di accesso* ai dati (access time) dipende da tre fattori fondamentali:

- Seek time

  La testina si sposta in senso radiale fino a raggiungere la traccia desiderata (impiega 1ms per tracce adiacenti, in generale tra 5 e 15ms in media).

- Latency time

  Quando è in funzione, i dischi ruotano ad una velocità tra 5400 e 10800 rpm. Occorre attendere, quindi, che il settore desiderato passi sotto la testina. Varia tra 2,7 e 5,4ms.

- Transfer time

  Tempo di lettura vero e proprio. In genere dipende dalla velocità di rotazione del disco e dalla densità di bit per cm lineare (della traccia). In genere si va da 5 a 20 MBps.

DISCHI MAGNETICI

Sono dotati di una grande capacità di memorizzazione. Fino a pochi anni fa i floppy disk rappresentavano il supporto mobile più usato.

Oggi sono stati largamente sostituiti da altri supporti quali, ad esempio, gli hard-disk esterni o i pen drive.

Un hard-disk, oggi, può raggiungere dimensioni di diversi TB, sono memorie più lente della memoria centrale, ma molto meno costose.

DISCHI OTTICI

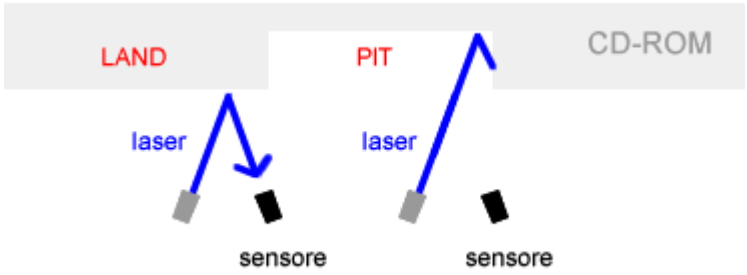
Un disco CD-ROM (memoria ottica) è un disco di policarbonato, in genere dal diametro di 12 cm, che ha la capacità di memorizzare 640MB, 700MB, 800MB o 870 MB (ben più rari).

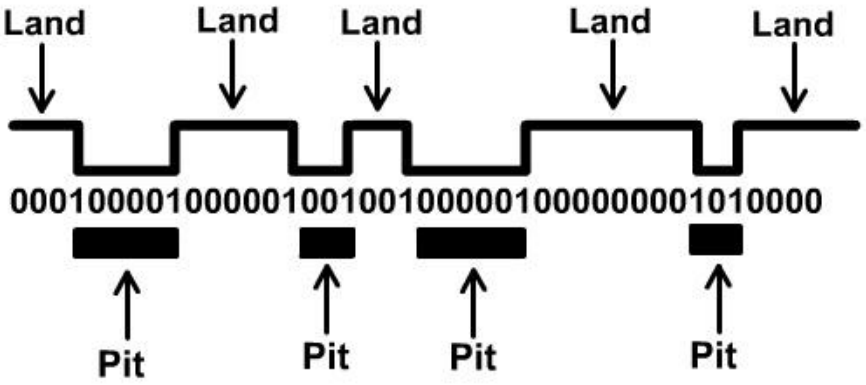
Sono stati oramai quasi totalmente sostituiti con i DVD, la cui capacità viene aumentata fino a 4,7 GB (se single layer - single side, ovvero lato unico - singolo strato) a 9,4 GB (se single layer double side, ovvero doppio lato - singolo strato) o a 17 GB (se double layer - double side, ovvero doppio lato – doppio strato).

I bit vengono impressi sotto forma di *pit* e di *land*. I pit sono dei solchi scavati nel substrato, le land sono zone non scavate del substrato. Il substrato viene realizzato vicino alla superficie e metallizzato, in modo da essere riflettente.

DISCHI OTTICI

Le transizioni *pit-land* e *land-pit*, nelle quali il sensore cambia di stato, vengono interpretate come un bit 1, mentre le aree piane presenti prima e dopo ogni transizione (nelle quali il sensore permane nello stato eccitato o non eccitato), vengono interpretate come 0 (o sequenze di zeri).





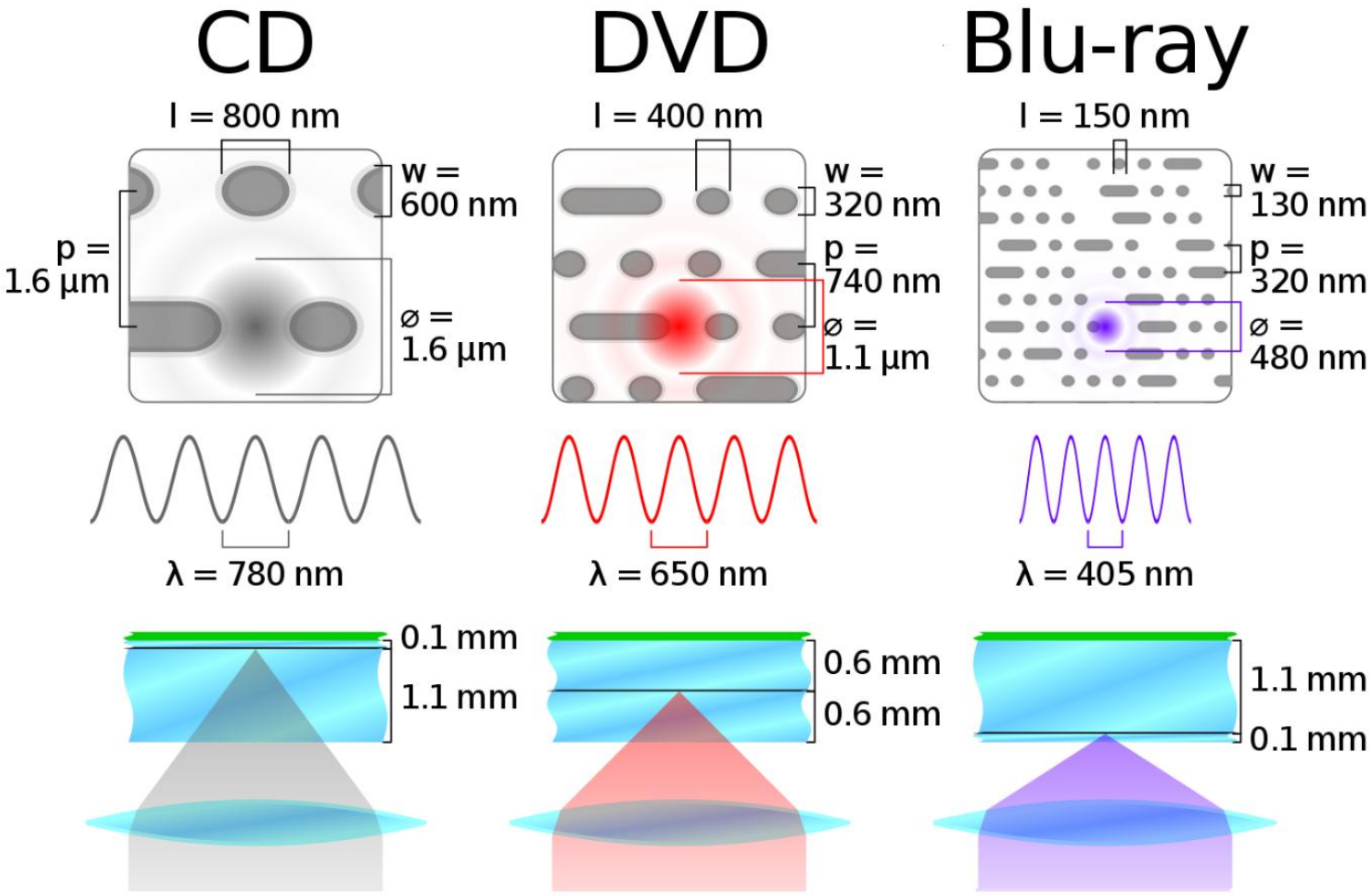
DISCHI OTTICI

Una recente evoluzione del DVD sono i Blu-ray, progettati dalla Sony per poter consentire la memorizzazione di video ad altissima risoluzione (4K).

Un supporto Blu-ray, oggi, può raggiungere capienze fino a 25GB, ma sono già in rilascio dischi che raggiungono fino a 50GB. La Sony lavora da tempo a varianti in grado di raggiungere perfino 200GB.

La differenza sostanziale consiste nell'utilizzo di un raggio laser a minor lunghezza d'onda (che assume una colorazione blu).

DISCHI OTTICI



FLASH MEMORY

È una tipologia di memoria *a stato solido*, di tipo non volatile, che per le sue prestazioni può anche essere usata come memoria a lettura-scrittura.

Diversamente dalle tecnologie precedenti, la tecnologia Flash ha reso possibile il salvataggio o la cancellazione di dati in un unico step, introducendo quindi un incredibile guadagno in velocità, e grazie alla non-volatilità è usata frequentemente nelle fotocamere digitali, nei lettori di musica portatili, nei cellulari,

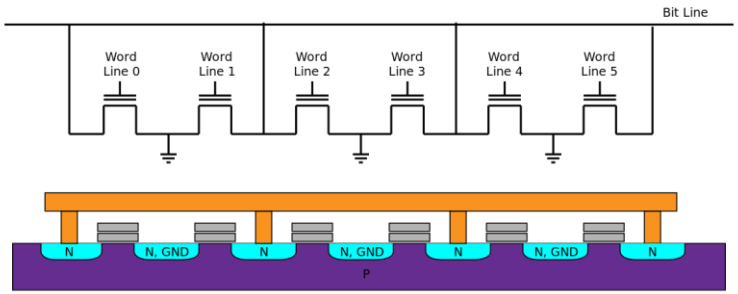
nei pendrive (chiavette), nei palmari, nei moderni computer portatili e in molti altri dispositivi che richiedono un'elevata portabilità e una buona capacità di memoria per il salvataggio dei dati.

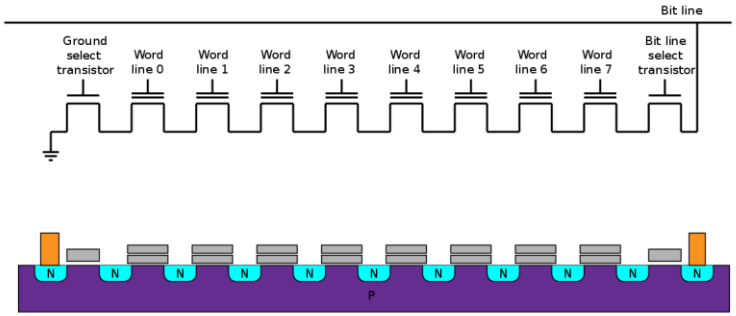
FLASH MEMORY

Esistono due tipologie differenti di memorie flash:

- NOR flash;
- NAND flash.

Nelle NOR, la bit line di ciascun array viene mantenuta al valore High (H), se tutte le celle sono poste al valore Low (L).

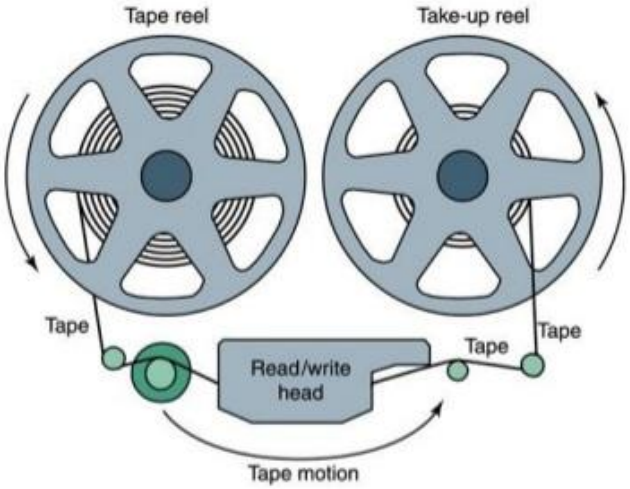




Nelle NAND, invece, la bit line di ciascun array può essere mantenuta al valore High (H), anche se una sola cella è posta al valore High (H).

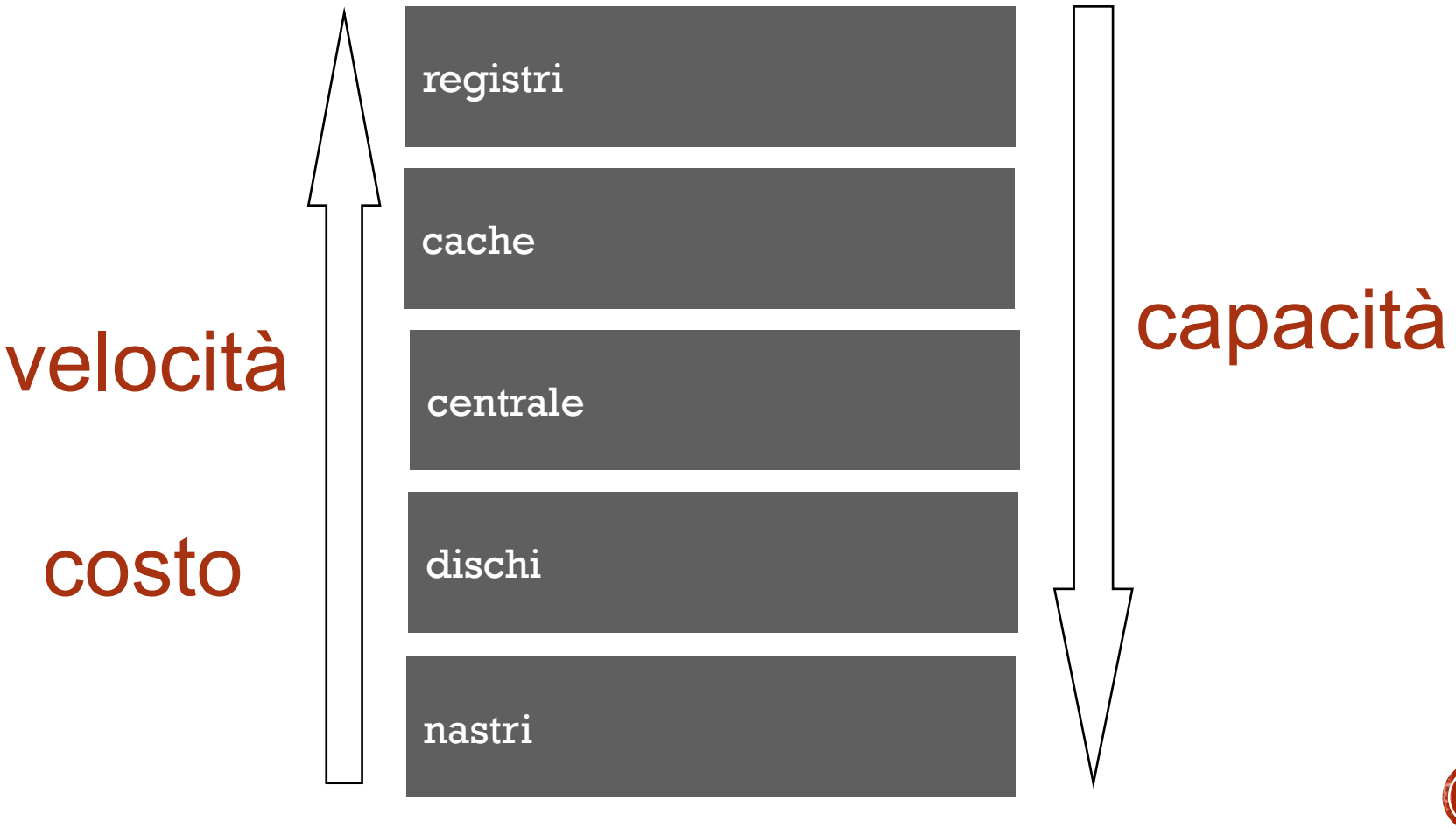
NASTRI MAGNETICI

Sono stati molto utilizzati nei primi calcolatori (anche destinati al mercato «home»), come supporto di memorizzazione per programmi e dati.



I nastri magnetici, oggi, vengono utilizzati esclusivamente come unità di *backup* per data center, in virtù della loro enorme capacità di memorizzazione ed al basso costo.

GERARCHIA DI MEMORIE



DISPOSITIVI DI INPUT/OUTPUT

I dispositivi di I/O, detti anche *periferiche*, permettono di realizzare l'interazione uomo/macchina. È importante notare che oltre alle periferiche di tipo tradizionale, ne esistono di varia natura per applicazioni specifiche (ad esempio in campo medico, geologico, ingegneristico, etc.).

La loro funzione principale è quella di consentire l'ingresso dei dati da elaborare, attraverso il bus, fino al processore e/o l'uscita delle informazioni restituite dalle elaborazioni, fino alla stampa su carta o a video.

Il controllo è eseguito da un software, detto "*driver*", che ne gestisce le funzionalità a basso livello (linguaggio macchina).

DISPOSITIVI DI INPUT/OUTPUT

Terminali

- Tastiera.
- Video.

Dispositivi di puntamento

- Mouse.
- Trackball.
- Penna ottica.
- Schermo sensibile (Touch Screen).

DISPOSITIVI DI INPUT/OUTPUT

Stampanti

Presentano varie tecnologie:

- Ad aghi.
- Ink jet.
- Laser.

Scanner, modem, plotter, scheda audio, etc.

UNITÀ DI INPUT

Tastiera

La tastiera (o keyboard) può essere ridotta, con 86 tasti, oppure espansa, con almeno 102 tasti. Si distinguono i seguenti elementi:

- la tastiera alfanumerica;
- i tasti funzione;
- Il tastierino numerico;
- i tasti di controllo:
  - ✓ STAMP, utilizzabile per stampare il contenuto dello schermo su carta;
  - ✓ CANC (o DEL), per cancellare il carattere su cui è collocato il puntatore;
  - ✓ (HOME) e (End), utilizzati nei WORD PROCESSOR per posizionarsi rapidamente all'inizio o alla fine della riga;
  - ✓ (PagUp) e (PagDn), per spostarsi rapidamente una videata in su o in giù;
  - ✓ TASTI FRECCIA, per spostarsi rapidamente in alto, in basso, a destra o a sinistra.

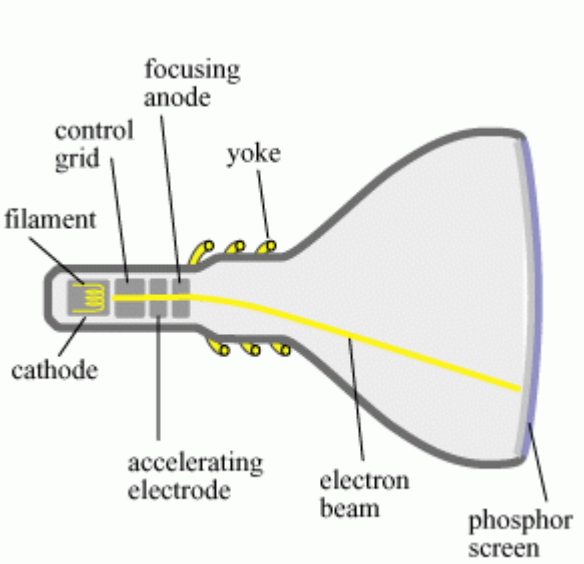
UNITÀ DI INPUT

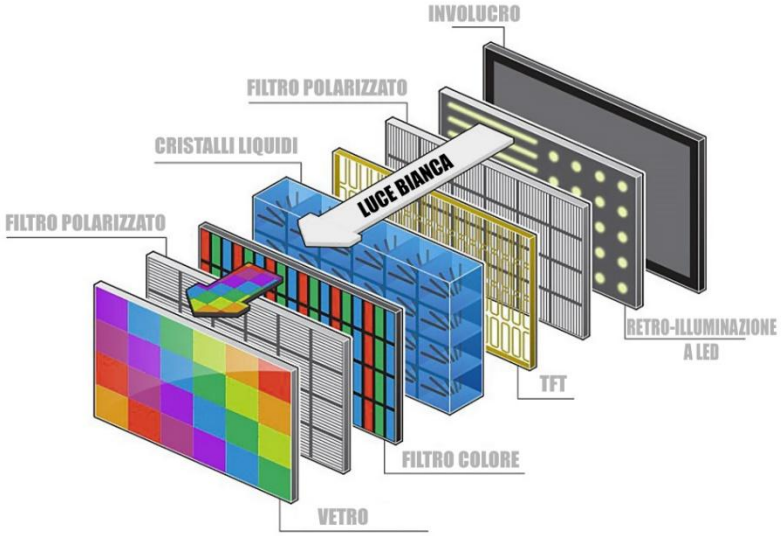
- **Mouse**.
- **Trackball**.
- **Touchpad** (sostituisce il mouse nei portatili).
- **Joystick**.
- **Penna ottica**.
- **Scanner**:
  - Manuali;
  - Piani.
- **Fotocamere** e **videocamere**.
- **Microfono**.

UNITÀ DI OUTPUT: MONITOR

Possono essere di tre tipi:

CRT (tubo a raggi catodici), utilizza la tecnologia simile a quella usata nei vecchi televisori.





LCD
(schermo a cristalli liquidi).

LED (schermo a cristalli liquidi con retroilluminazione a led).

UNITÀ DI OUTPUT: MONITOR

Parametri caratteristici di un monitor sono:

Pollici

unità di misura del monitor (calcolata sulla diagonale dello schermo). Un pollice equivale a 2,54cm. Il rapporto base/altezza può valore 4/3 o 16/9.

Pixel

Il monitor può essere pensato come una griglia composta da migliaia di quadratini detti *pixel*. Maggiore è il numero di pixel, più definita è l'immagine ottenuta nello schermo.

Risoluzione del monitor

Si esprime sempre in pixel e indica il numero di punti in orizzontale e in verticale sullo schermo (es. 1024 x 768). I formati standard più usati, oggi, prevedono le seguenti risoluzioni:

HD (o HDTV): 1280 x 720 (ovvero 921.600)

Full HD: 1920 x 1080 (ovvero 2.073.600 px)

Ultra HD: 3840 x 2160 (ovvero 8.294.400 px)

4K (Digital Cinema): 4096 x 2160 (ovvero 8.847.360 px)

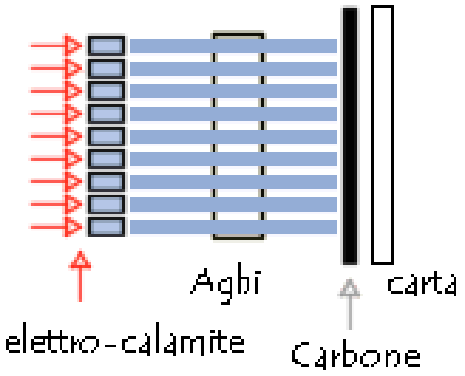
UNITÀ DI OUTPUT: STAMPANTI

Esistono diverse tecnologie per le stampanti:

- A impatto
  - ✓ aghi.
- A non impatto
  - ✓ getto di inchiostro.
  - ✓ laser.

Le **stampanti ad ago** usano una matrice in genere a 9 o a 24 aghi, posti in verticale, con la quale scrivono i caratteri per singola colonna.

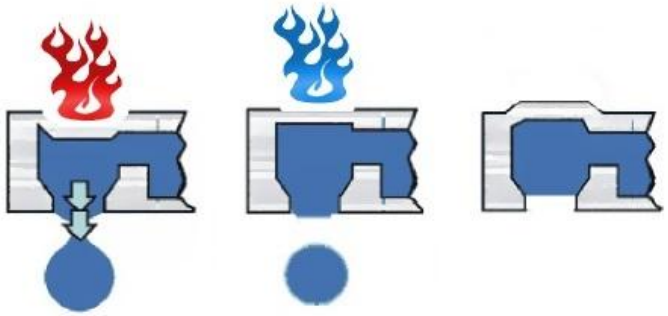
Tra il foglio e gli aghi, viene posizionato un nastro preinchiostrato. Quelle a 24 aghi arrivano ad una risoluzione di 216 ppp.



UNITÀ DI OUTPUT: STAMPANTI

Le stampanti a **getto d'inchiostro** possono essere a *testina termica*:

ogni ugelli ha un elemento riscaldante (un resistore). Il calore fa uscire una goccia di inchiostro, mentre il raffreddamento fa rientrare l'inchiostro.



Vi sono poi quelle a *testina piezoelettrica (a bolla di inchiostro)*:

ogni ugello ha un quarzo piezoelettrico che, se eccitato sulla sua frequenza di risonanza, si deforma facendo prima uscire la goccia di inchiostro a pressione e poi staccandola per risucchio.

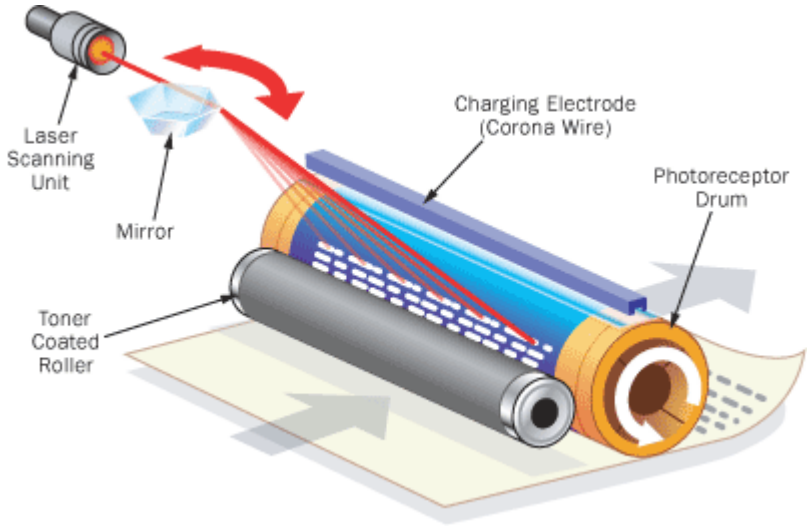
UNITÀ DI OUTPUT: STAMPANTI

Le **stampanti laser**, infine, fanno uso del toner, una polvere finissima contenente carbone e ossido di ferro.

Un rullo fotosensibile viene caricato negativamente. Attraverso un laser si neutralizza tale carica dove il toner non deve attaccare. Il toner, caricato positivamente, verrà attratto dal rullo.

Un foglio di carta viene caricato negativamente (con una intensità maggiore) in modo da catturare il toner.

Il foglio viene fatto passare nel cosiddetto «forno di cottura» fissandone il toner. Quello in eccesso verrà spazzolato via, prima dell'uscita.



UNITÀ DI OUTPUT

Altre periferiche di output sono:

Plotter

Simile al tecnigrafo, viene utilizzato principalmente per stampare su modulo continuo disegni di grandi dimensioni.

Microfilm

Consentono l'archiviazione di grandi quantità di dati sotto forma di immagine.

Altoparlanti

Sono collegati alla scheda audio.

LE PORTE

Seriale
Trasmette un bit alla volta (ha 9 o 25 pin).

Parallela
8 bit alla volta, 25 pin, LPT (**L**ine **P**rin**T**er).

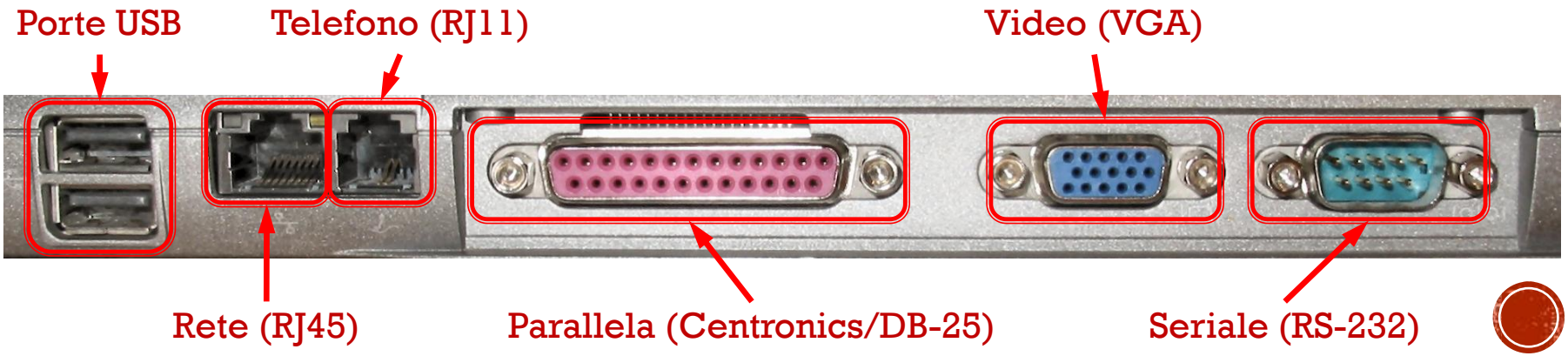
SCSI
8, 16, 32 bit.

WiFi, Bluetooth, HDMI, …

USB
Universal serial bus.

FireWire
Detta anche IEEE 1394.

IRDA
infrared data association.



IL SISTEMA OPERATIVO E GLI APPLICATIVI SOFTWARE

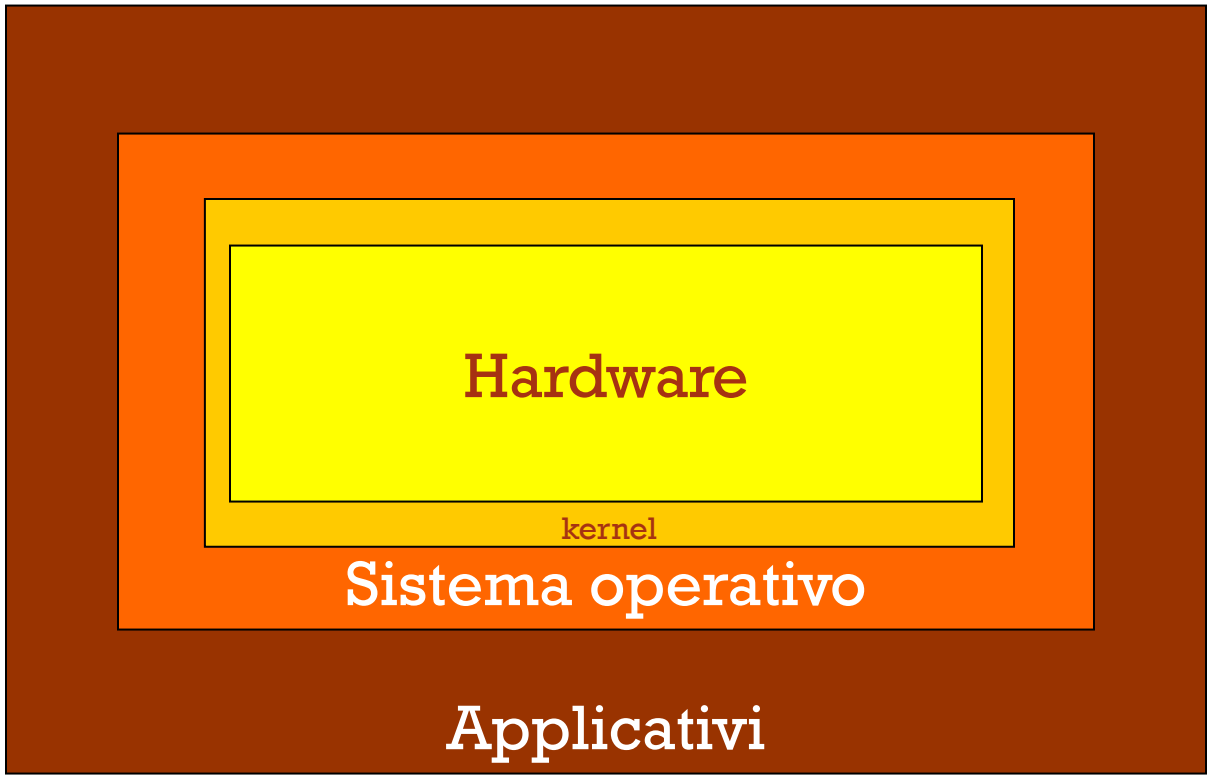
IL SOFTWARE

Insieme di programmi la cui esecuzione permette di creare l'ambiente virtuale con cui l'utente si trova ad interagire.

I programmi eseguibili sono scritti in *linguaggio macchina*. Per linguaggio macchina si intende quel linguaggio di programmazione a basso livello, le cui singole istruzioni sono direttamente eseguibili dalla macchina Hardware e sono formate da stringhe binarie o al più dalla loro «astrazione» attraverso stringhe.

Talvolta si tende a confonderlo con il linguaggio Assembly, che in realtà, pur essendo un linguaggio a basso livello, rimane pur sempre un linguaggio simile a quelli evoluti (C, C++, Java, etc.) che, come vedremo, *richiedono un processo di traduzione*.

IL SOFTWARE



IL SISTEMA OPERATIVO

È un insieme di programmi che gestiscono e coordinano le varie risorse dell'elaboratore, facilitandone l'uso.

Esso costituisce l'interfaccia tra la macchina hardware e l'utente, fornendo una sorta di visione astratta della stessa.

In definitiva, il Sistema Operativo è un insieme di *moduli software* che virtualizzano l'hardware sottostante, creandone una *istanza astratta* più semplice da utilizzare rispetto a quella *fisica*.

IL SISTEMA OPERATIVO

Il Sistema Operativo (SO), principalmente, presenta le seguenti attività funzionali:

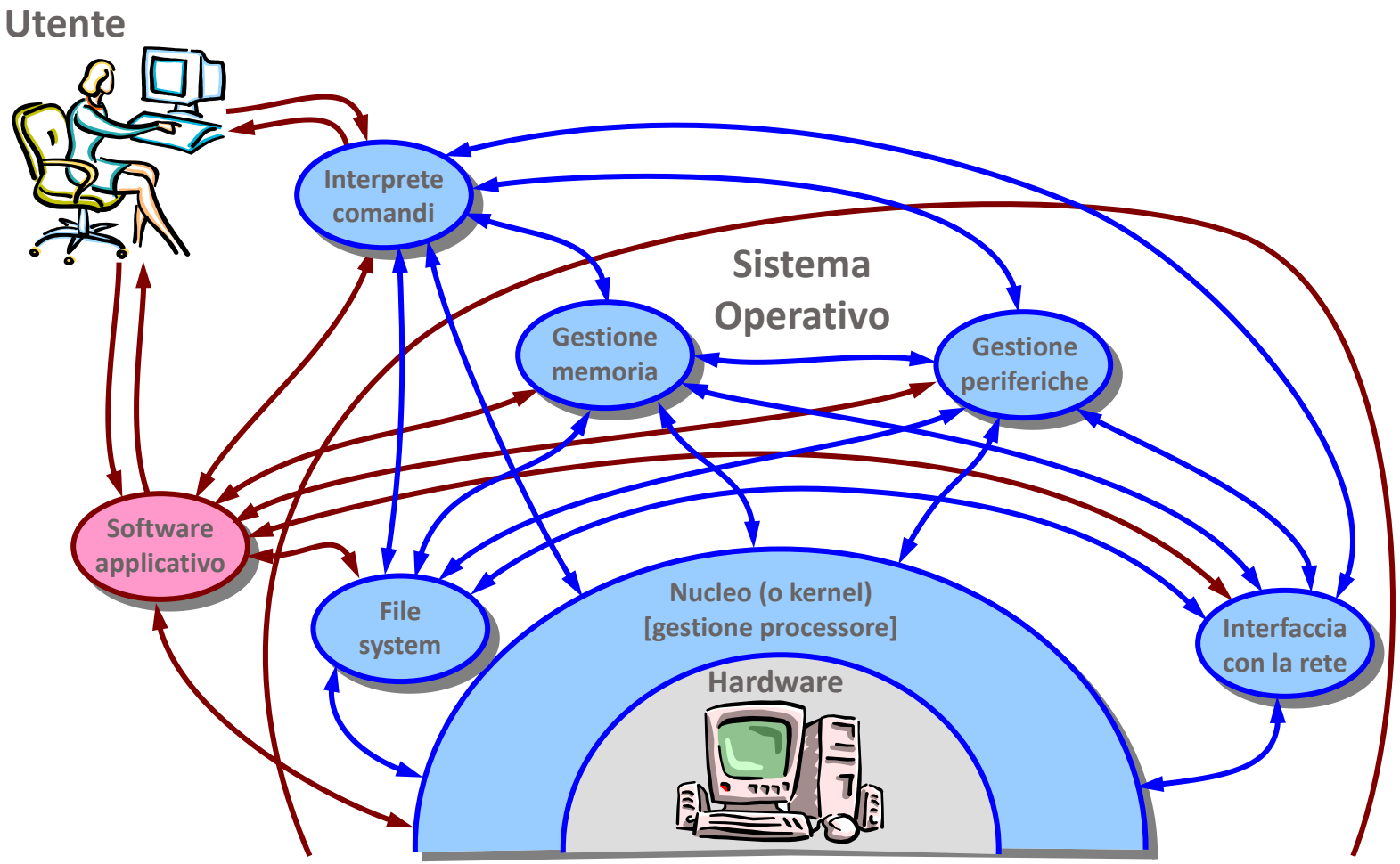
- gestione
  - ✓ del processore;
  - ✓ della memoria principale;
  - ✓ delle risorse;
  - ✓ delle informazioni in memoria secondaria (file system);
- configurazione della macchina all'accensione
- macchina estesa
  - ✓ costituisce la base sulla quale scrivere programmi;
  - ✓ mostra una macchina estesa più semplice da programmare rispetto all'hardware sottostante.

IL SISTEMA OPERATIVO

Se si vuole entrare in dettaglio, le funzioni di un SO sono:

- esecuzione di programmi
  Li carica, insieme ai dati, in memoria centrale, inizializza le periferiche di I/O, prepara e gestisce risorse quali la rete, etc.
- accesso ai dispositivi di I/O
  Gestisce i segnali nel bus, consentendo all'utente di farne un uso astratto.
- archiviazione di dati e/o programmi
  Fornisce una organizzazione logica in file, cartelle (directory), gestendo il flusso di basso livello, nel bus, con le periferiche di I/O.
- controllo di accesso
  Meccanismi di protezione e condivisione delle risorse tra utenti o processi, con relativa gestione dei conflitti.
- Gestione dei malfunzionamenti
  Rilevare e, dove possibile risolvere, malfunzionamenti all'hardware causati da operazioni errate compiute dai programmi.

IL SISTEMA OPERATIVO



IL SISTEMA OPERATIVO

Classificazione dei sistemi operativi:

- *mono-utente (mono-user)*

  il sistema può essere utilizzato da un solo utente per volta
- *multi-utente (multi-user)*

  consente l'interazione contemporanea a più utenti (in questo caso il SO fornisce un sistema astratto dedicato a ciascuno di loro)
- *mono-programmato (mono-task)*

  consente l'esecuzione di un solo programma per volta
- *multi-programmato (multi-task)*

  come visto, i programmi vengono eseguiti «contemporaneamente» grazie al time sharing (ad ogni task, o processo, viene assegnato un certo numero di clock per l'esecuzione da parte del processore, rimanendo in standby mentre accade la stessa cosa per gli altri task)
- *distribuito (sistemi operativi di rete)*

  più elaboratori, locali o remoti, vengono gestiti come un unico sistema al fine di compiere operazioni complesse (calcolo distribuito) o gestire un elevato numero di accessi (cluster in datacenter)

LA FASE DI BOOTSTRAP

In tutti i calcolatori, al momento dell'accensione vengono eseguite una serie di operazioni che si completano con il caricamento in memoria centrale del kernel del SO, prelevandolo dalla memoria secondaria.

Tale fase prende il nome di *bootstrap*. Se nella macchina sono presenti più SO, solitamente è possibile selezionare quale far partire attraverso il boot manager, per poi passare le indicazioni al *boot loader*.

Il *kernel* è la parte del Sistema Operativo che agisce a stretto contatto con l'hardware.

LE GUI

Windows

Nasce intorno alla fine degli anni ottanta, come ambiente operativo grafico per computer con Sistema Operativo MS-DOS.

Prima di allora, si disponeva solo di shell testuali attraverso le quali ogni operazione necessitava di digitare opportuni comandi, con relativi parametri (ove richiesti), con un prompt del tipo c:\>

I comandi erano classificati in comandi per la manipolazione di dischi e files, comandi di interazione con le risorse hardware e comandi di utilità per lo sviluppo di programmi.

LE GUI

Windows

L'interfacce grafica, detta GUI (Graphical User Interface), ha reso l'utilizzo dei computer User Friendly.

Alla grafica, si è aggiunta una importantissima categoria di periferiche hardware, ovvero i sistemi di puntamento grazie ai quali selezionare oggetti grafici visualizzati sullo schermo ed interagire con loro.

Ciò che compare a video può essere paragonato ad una scrivania (desktop) sulla quale sono presenti strumenti ed oggetti rappresentati mediante disegni (le icone) che ricordano in modo diretto la funzionalità implicita.

LE GUI

I documenti (files) possono essere raccolti in cartelle (le directory) dando loro dei nomi significativi come si farebbe con una etichetta cartacea attaccata su di una carpetta reale.

Quando si manda in esecuzione un programma, la relativa istanza in esecuzione avrà un ambiente grafico dedicato (la finestra, o windows) dentro il quale si svolgerà l'interazione completa con l'utente. Ogni finestra è dotata di pulsanti per ridurre ad icona, ridimensionare o chiuderla.

Chiudere una finestra equivale ad interrompere l'esecuzione del relativo programma, con il conseguente rilascio delle risorse utilizzate.

GESTIONE FILES

Per assegnare un nome ai files (che di fatto sono una sorta di contenitori logici per i bit che costituiscono dati o programmi), nella maggior parte dei sistemi operativi viene adottata la seguente convenzione:

nome_file.*estensione*

Nella directory dentro la quale è memorizzato il file, *nome_file* deve essere un identificativo univoco. L'estensione, invece, è una parte facoltativa costituita da un un testo lungo da 1 a 4 caratteri che identifica il contenuto del file stesso.

Esempio:

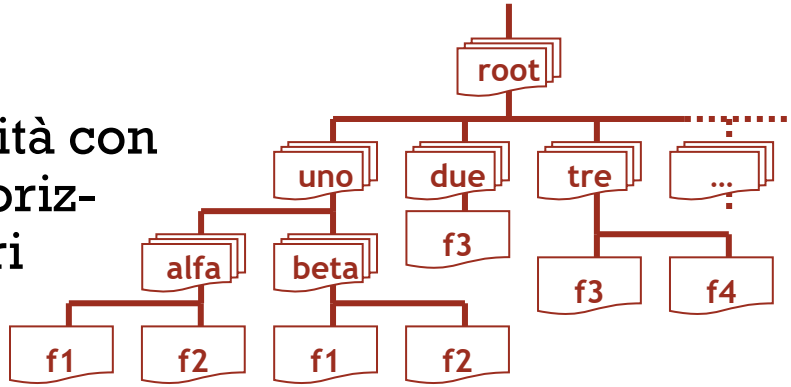
lettera.docx

La directory principale è detta anche root e viene creata in automatico al momento della formattazione del disco rigido.

IL FILE SYSTEM

È il meccanismo attraverso il quale vengono organizzati i files all'interno di una memoria di massa. Per essere più formali, è l'insieme dei tipi di dati astratti utilizzati per la scrittura, l'organizzazione gerarchica, la modifica e la lettura dei file sulle memorie di massa.

Il file system organizza la modalità con la quale ciascun file viene memorizzato a livello fisico sui vari settori della memoria di massa.



Ciascun file system, poi, dispone di tabelle per associare ad ogni nome_file il corrispondente collegamento fisico al file stesso, ovvero l'indice correlato in una tabella di allocazione dei file. Non tutti i file system hanno bisogno di una memoria fisica.

SOFTWARE APPLICATIVO

Rappresenta l'insieme di programmi in grado di soddisfare le più svariate esigenze degli utilizzatori.

Le più diffuse classi di programmi applicativi sono:

- Elaboratori di testi (Word processors);
- Fogli elettronici (Spreadsheets);
- Basi di dati (Databases);
- Programmi di grafica ed animazione;
- Programmi multimediali;
- Traduttori (interpreti e compilatori);
- Contabilità aziendale;
- Intrattenimento e gioco;
- Di presentazione
- Desktop publishing (per l'editoria)

Il software applicativo può essere di due tipi:

- General purpose.
- Orientato alla risoluzione di uno specifico problema.

ELABORATORI DI TESTI

Editors di caratteri ASCII

- BloccoNote di Windows.
- Emacs.

Word processors

- Microsoft Word.
- Word-pad di Windows.
- Writer (incluso in Open Office).
- …

Sistemi di impaginazione basati su testo formattato (il documento viene impaginato in seguito ad una traduzione).

- TeX (o la sua variante LaTeX).
- HTML.

FOGLI ELETTRONICI

Consentono di:

- effettuare elaborazioni matematiche e statistiche su dati raccolti in forma tabellare;
- realizzare grafici di vario tipo;
- filtrare i dati da elaborare/visualizzare;
- Formattare graficamente i dati gestiti.

Principali programmi commerciali:

- Lotus;
- Microsoft Excel;
- Calc (incluso in Open Office);
- …

BASI DI DATI

Le basi di dati servono a raccogliere ed a gestire in maniera organizzata, enormi quantità di dati. Più avanti, vedremo meglio il concetto di dato ed i benefici apportati da una loro gestione attraverso un DBMS.

Basi di dati sul modello relazionale:

- Dbase;
- Microsoft Access;
- Oracle;
- SQL Server;
- …

GRAFICA ED ANIMAZIONE

Consentono di creare o di elaborare immagini ed animazioni in maniera interattiva (ad esempio "foto-ritocco").

Alcuni esempi di software appartenente a tale categoria:

- Photoshop.
- CorelDraw.
- Paint shop.
- Blender (di tipo Open Source).
- …

PROGRAMMI MULTIMEDIALI

L'interazione multimediale è la mescolanza di:

- suoni;
- immagini;
- video;
- iper-testi.

L'uso di strumenti multimediali si è diffuso in modo significativo con Internet. Programmi per la creazione di siti Web:

- Flash;
- DreamWeaver;
- Webmatrix (gratuito);
- …

TRADUTTORI

Quando un programmatore realizza un software utilizzando un linguaggio di programmazione di alto livello, di fatto realizza un file di testo chiamato *codice sorgente* (o semplicemente *sorgente*), seguendo le regole sintattiche e semantiche previste dal linguaggio stesso.

Un programma viene progettato a partire da un particolare problema che si intende risolvere, trasformandolo in un algoritmo (eventualmente ricorrendo ad un diagramma di flusso).

Di fatto, un file sorgente conterrà delle dichiarazioni di variabili e/o costanti (e loro inizializzazioni), funzioni, istruzioni e strutture di controllo del flusso di esecuzione, il tutto in accordo con il paradigma di programmazione scelto (dal quale dipende, in conseguenza, il linguaggio di programmazione).

TRADUTTORI

Per poter essere eseguito, il file sorgente dovrà essere tradotto in *codice oggetto* (o *file oggetto*), scritto in linguaggio macchina (quindi in binario).

Il procedimento di traduzione, (in un determinato caso anche attraverso il link alle varie librerie software utilizzate), arriverà a produrre un file eseguibile (in windows con estensione .exe) che potrà essere avviato con un doppio clic sulla relativa icona, in modo da fruire del programma realizzato.

Il file eseguibile, a quel punto, potrà essere distribuito in altri computer con caratteristiche simili a quello per il quale è stato sviluppato.

TRADUTTORI

La distribuzione del software proprietario (closed source), tuttavia, è soggetta a particolari forme contrattuali che prendono il nome di *licenze d'uso*.

Esistono varie forme di licenza. Giusto per citare alcune delle più importanti:

- EULA
  End-User License Agreement (EULA) è un contratto attivato tra l'autore del software e l'utente finale che assegna la licenza d'uso del software secondo i termini stabiliti.
- SHAREWARE
  Il software viene distribuito liberamente e può essere utilizzato gratuitamente in versione di prova o per un periodo di tempo limitato, o con funzionalità ridotte. È un modo molto diffuso tra chi desidera valutare se procedere con una EULA.
- FREEWARE
  Il software può essere distribuito gratuitamente secondo le disposizioni dell'autore.

TRADUTTORI

I software possono essere distribuiti anche in modalità open source, ovvero attraverso il rilascio non solo del file oggetto ma anche dei files sorgenti.

La filosofia dietro l'open source si è estesa ad altri movimenti che puntano alla diffusione in modalità open di contenuti editoriali e perfino dell'hardware.

vEyes è nato con l'intento di estendere alle tecnologie assistive ed agli strumenti di diagnostica tale filosofia.

Anche per la distribuzione del software libero esistono delle licenze. Tra tutte possiamo citare la GNU General Public License (GPL), la GNU Lesser General Public License (LGPL) o le licenze BSD.

TRADUTTORI

A seconda del linguaggio di alto livello utilizzato tra i tanti disponibili (ad esempio C++, Java, C#, Visual Basic, etc.) si possono avere due tipi differenti di traduttori:

- Interpreti

  ogni istruzione del file sorgente viene tradotta ed eseguita in tempo reale, procedendo istruzione per istruzione

- Compilatori

  La traduzione del file sorgente avviene per l'intero programma, generando il file oggetto che è può essere così eseguito.

LE RETI DI CALCOLATORI ED INTERNET

RETE DI CALCOLATORI

É un insieme di calcolatori, collegati tra loro mediante una *rete di comunicazione*, che possono interagire tra loro, condividere informazioni e risorse, ma che possiedono comunque una indipendenza nella funzionalità, nel momento in cui intendano operare in modalità stand-alone.

Per rete di comunicazione si intende una infrastruttura formata da *nodi* (computer, server, banche dati, etc.) connessi tra loro attraverso un *sistema di interconnessioni* (canali fisico) che scambiano dati grazie ad un *protocollo di comunicazione*.

RETE DI CALCOLATORI

In definitiva, attraverso una rete informatica, è possibile:

- condividere in modo razionale risorse costose
  ad esempio, non è conveniente dotare ogni computer di una stampante multifunzione;
- condividere/scambiare dati e/o informazioni
  ad esempio un database al quale molti utenti, da remoto, devono poter accedere per leggere/scrivere;
- creare team di lavoro geograficamente distribuiti
  (si pensi alla didattica effettuata da remoto o allo smart working);
- fornire servizi.
  si pensi ad un servizio di booking, o di home banking, giusto per citarne alcuni esempi.

RETE DI CALCOLATORI

TIPOLOGIE DI RETI

È possibile identificare due macro-tipologie di reti di calcolatori:

- reti locali

  collegano elaboratori in un'area geografica circoscritta (stesso edificio o edifici attigui), distribuiti in un contesto privato;

- reti geografiche

  collegano elaboratori in località remote, distribuiti in un contesto geografico ampio (urbano e/o extraurbano).

TIPOLOGIE DI RETI

Entrando in dettaglio, possiamo avere dunque:

- LAN (Local Area Network)

  rete composta da nodi che risiedono, senza sconfinare, in un'area privata (solitamente uno stesso ambiente di lavoro);
- MAN (Metropolitan Area Network)

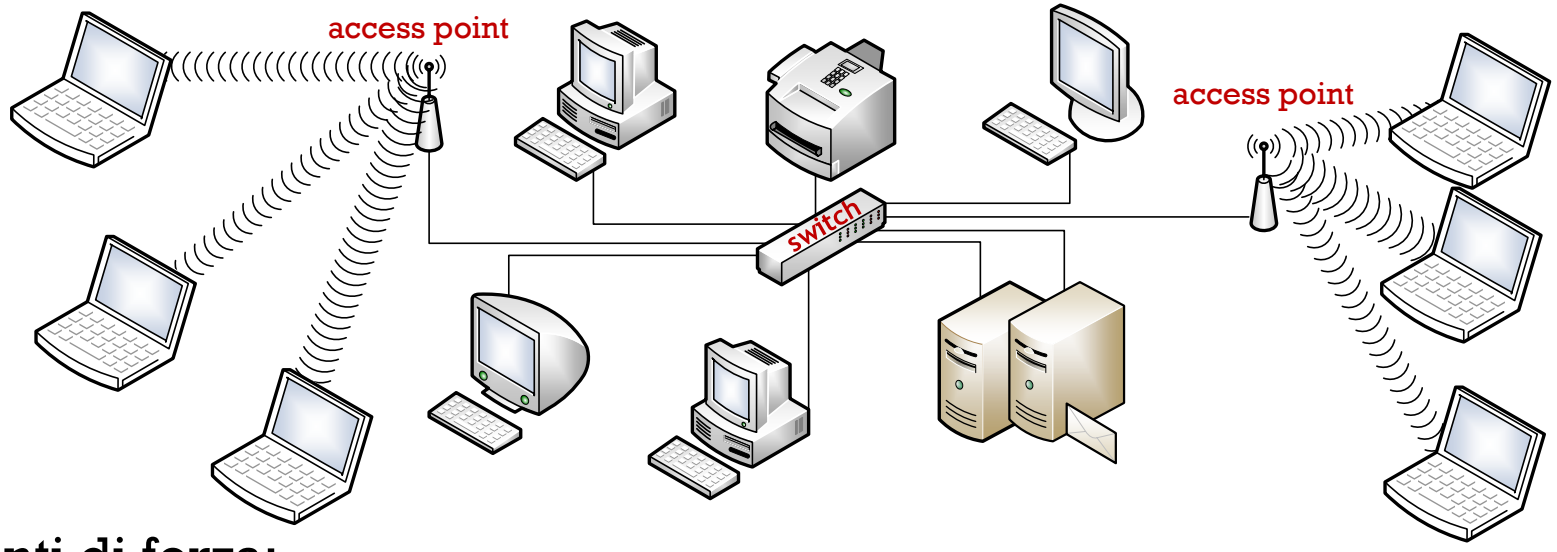
  rete composta da nodi che risiedono all'interno di una stessa area urbana;
- WAN (Wide Area Network)

  rete composta da nodi che sconfinano oltre l'ambiente urbano e perfino oltre uno stesso stato (rimanendo dentro un continente).
- GAN (Global Area Network)

  rete composta da nodi dislocati in tutti i continenti.

RETI LOCALI

Sono reti che coprono distanze circoscritte ad un contesto privato. Solitamente fanno uso quasi esclusivo di connessioni via cavo (Ethernet), sebbene sia sempre più diffusa la configurazione mista (cablata/wireless):



Punti di forza:

- solitamente più veloci e sicure di una rete geografica;

RETI GEOGRAFICHE

Sono reti di grosse dimensioni che coprono lunghe distanze (come detto, anche intercontinentali).

Il sistema di interconnessioni fa uso di canali (wired e/o wireless):

- linee telefoniche,
- satelliti,
- fibre ottiche.

Si definisce *larghezza di banda* la capacità di trasmissione di una rete misurata in bit al secondo (bps, multiplo più usato Mbps).

Punti di criticità:

- solitamente più lente di una rete LAN;
- a bassa affidabilità (problemi di sicurezza).

RETI GEOGRAFICHE

Punti di forza:

- permettono la comunicazione tra utenti in locazioni fisiche distinte (scambio di messaggi e dati) anche parecchio distanti tra loro.

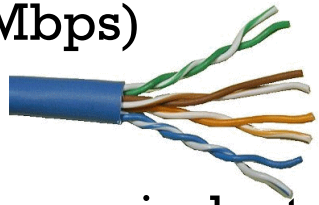
Contesti applicativi:

- comunicazioni in ambito di ricerca;
- utilizzo di basi di dati in locazioni remote;
- lavoro cooperativo;
- possibilità di svolgere attività di lavoro a casa,
- accesso a informazioni di varia natura (Internet).

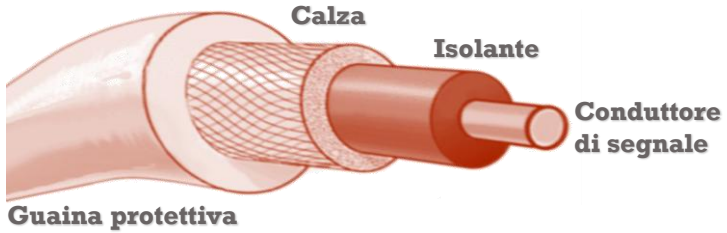
GLOSSARIO

Nodo: singolo elemento di rete, può essere il semplice PC, un server, una stampante di rete, etc.

Doppino telefonico: cavo costituito da due fili di rame intrecciati. Non è il più veloce ma il più diffuso. Garantisce la trasmissione nei due sensi (velocità 1÷200 Mbps)

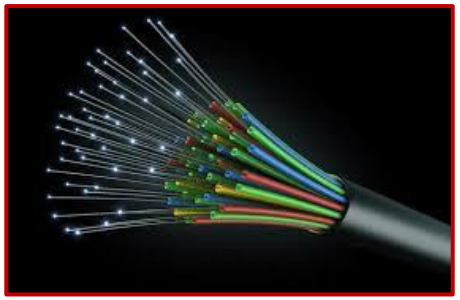


Cavo coassiale: filo di rame ricoperto da un isolante PVC o teflon, avvolto da una calza schermante (gabbia di Faraday). Usato nelle reti locali (velocità 500 Mbps).

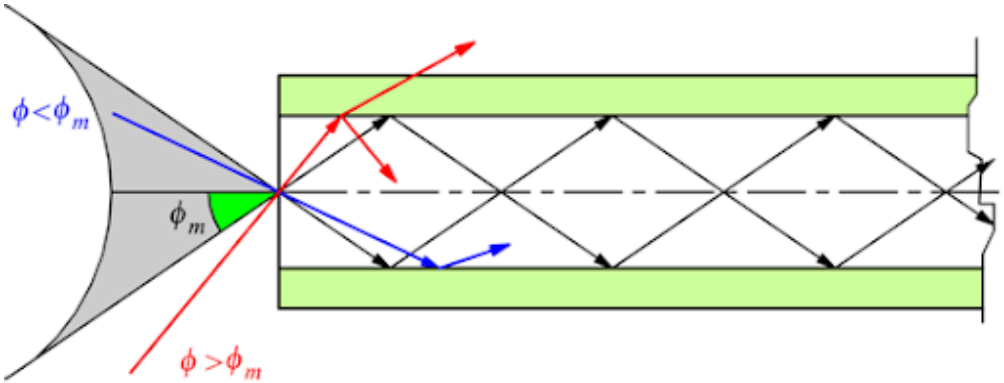


GLOSSARIO

Fibre ottiche:
indica dei cavi al cui interno è presente uno o più fili vetrosi (fibre di vetro), protetti da particolari guaine, che sono in grado di condurre la luce a grandi distanze (velocità 10 Gbps).



Un emettitore di laser invia il segnale luminoso. All'altra estremità è presente un ricevitore che converte il segnale da luminoso in elettrico. Se l'angolo incidente con il quale il laser colpisce l'esterno della fibra è inferiore ad un certo valore, si ha la riflessione interna totale (assenza di rifrazione):



IL MODEM TRADIZIONALE

La rete telefonica, tipicamente, trasmette i dati in formato analogico. Come fanno due computer all'interno di una rete geografica a comunicare?

Il dispositivo da utilizzare è il **Modem** (**Mo**dulatore-**Dem**odulatore) al quale è affidato il compito di modulare (da digitale ad analogico) o demodulare (da analogico a digitale) il segnale con i dati scambiati tra il computer (posto a monte del Modem) e la rete esterna (posta a valle del Modem).

La velocità di trasmissione è espressa in baud. Equivale al numero di segnali modulati al secondo (ovvero il numero di parole da 8 bit).

COMUNICAZIONE TRA PIÙ MACCHINE

Il collegamento tra due o più nodi più nodi della rete, attraverso canali fisici wired o wireless, può essere di due differenti tipologie:

- *dedicato*, ovvero tramite una linea privata permanente, sempre presente a prescindere che ci si scambi o meno dei ;
- *commutato*, ovvero il collegamento tra i nodi non è fisso ma si realizza, in modo temporaneo, solo al momento dell'effettiva trasmissione dei dati.

TIPI DI COLLEGAMENTI

PSDN (**P**ublic **S**witched **D**ata **N**etwork) rete pubblica di dati commutata. Rappresenta la linea telefonica analogica.

- Linea a commutazione di circuito.

ISDN (**I**ntegrated **S**ervice **D**igital **N**etwork) rete digitale integrata. Identifica la linea telefonica digitale.

ADSL (**A**symmetrical **D**igital **S**ubscriver **L**ine) tecnologia di compressione dati che consente la trasmissione ad alta velocità utilizzando le normali linee telefoniche.

IL MODEM ISDN

Mentre i modem di vecchia generazione e quelli ADSL sono veri e propri **MO**dulatori/**DEM**odulatori, quelli ISDN sono chiamati così solo per semplificazione, in quanto non usano frequenze sul doppino telefonico, ma veri e propri segnali digitali discreti.

La definizione corretta è TA (Terminal adapter) ISDN. Sono utilizzabili solo se in possesso della linea telefonica specifica per questo tipo di connessione.

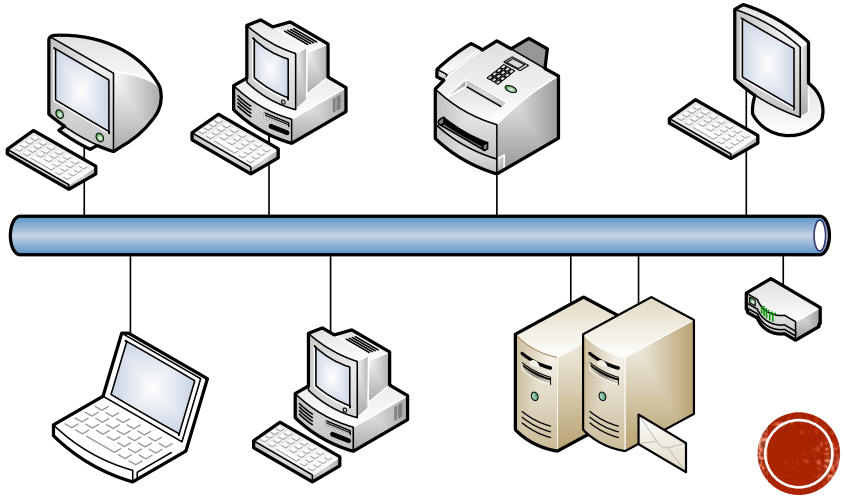
TIPOLOGIE DI CONNESSIONI

Le connessioni tra i nodi possono essere di due tipi:

- reti *multipunto* (broadcast);
- reti *punto a punto*.

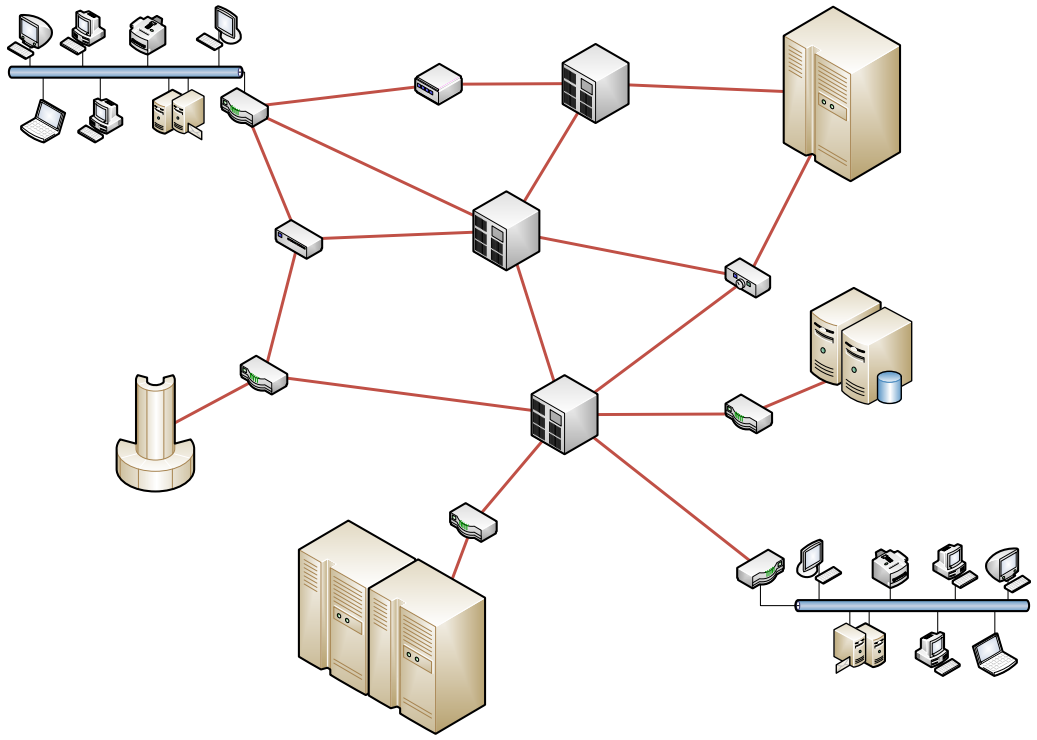
Nelle reti multipunto il sistema di comunicazione viene condiviso da tutti i nodi. Ad ogni nodo è associato un indirizzo di rete.

Un pacchetto raggiunge tutti i nodi ma verrà trattenuto (ed elaborato) solo dal nodo che possiede lo stesso indirizzo di rete indicato nel pacchetto come destinatario.



TIPOLOGIE DI CONNESSIONI

Nelle reti punto a punto, invece, il sistema di comunicazione e formato da un insieme di coppie di nodi tra loro connesse.



Un pacchetto inviato da un nodo, deve seguire un percorso che passi attraverso più nodi i quali provvederanno ad instradarlo fino al nodo di destinazione.

PROTOCOLLI DI COMUNICAZIONE

Se la parte fisica (nodi e sistema di interconnessione) di una rete consente la comunicazione tra computer, affinché la comunicazione avvenga in modo corretto è necessario definire un *protocollo di comunicazione*.

Il protocollo di comunicazione specifica le regole con le quali i diversi dispositivi interagiscono tra loro. Ad esempio:

- stabilisce in modo preciso come associare un nome logico e un indirizzo fisico ai nodi della rete;
- stabilisce come procedere in caso di errore, perdita di pacchetti o di ritardo durante la comunicazione;
- etc.

INTERNET

È nata negli USA, durante il periodo della guerra fredda, quale sistema di difesa e di controspionaggio.

Finita l'epoca della guerra fredda, intorno alla fine degli anni '60 venne estesa ad ARPANET, una rete che collegava tra loro 4 nodi formati da altrettante università americane.

In pochi anni, ARPANET estese i nodi anche oltre oceano, fino a diventare, a partire dalla fine degli anni '80 quello è oggi Internet (detta anche la «rete delle reti»).

La svolta la si deve alla nascita del protocollo HTTP (HyperText Transfer Protocol) nel 1991 (grazie al quale sono nati gli ipertesti) ed a seguire del servizio www (che permette di navigare documenti collegati tra loro da link).

INTERNET

Ad internet, oggi, sono collegati, giusto per citarne alcuni:

- Organizzazioni Internazionali
  L'ONU, la FAO, la Banca Mondiale
- Organismi politici
  Casa Bianca, Parlamenti e ministeri di varie nazioni.
- Biblioteche
- Mass Media di varia natura
  Giornali, riviste, agenzie di stampa, reti televisive.
- Associazioni scientifiche e professionali
- Aziende
- Enti pubblici
- Enti commerciali (commercio elettronico)
- Società di servizi

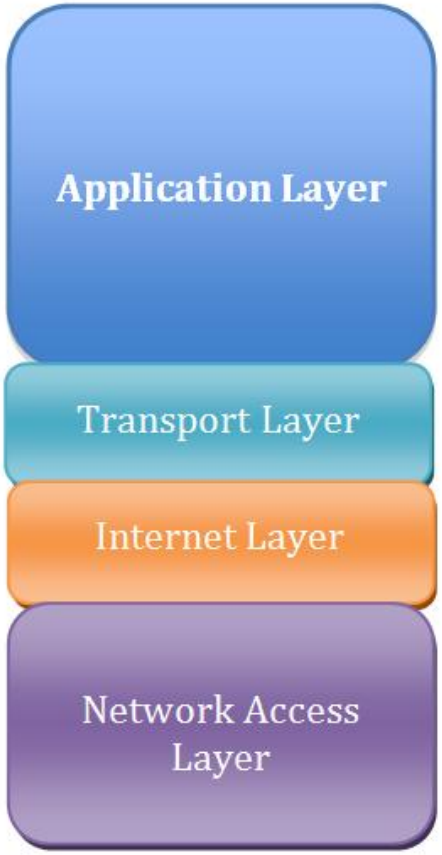
INTERNET: IL MODELLO TCP/IP

Questo modello prevede 4 livelli, ciascuno con diversi compiti:

- Accesso alla rete
  tale livello, nella pratica, fa uso di Ethernet (cavo) e IEEE 802.11 (radio) e definisce lo scambio di dati fra un sistema finale e la rete a cui è collegato, specificando come organizzare i dati in frame e come trasmetterli sulla rete;
- IP (Internet Protocol)
  definisce lo scambio di dati tra sistemi che non appartengono alla stessa rete attraverso procedure per attraversare reti multiple interconnesse. Specifica il formato dei pacchetti inviati attraverso la rete e i meccanismi utilizzati per farli transitare dal nodo sorgente, attraverso uno o più router, verso il destinatario;
- Trasporto (host to host) – TCP (Transmission Control Protocol)
  garantisce che tutti i pacchetti giungano a destinazione e definisce i protocolli che consentono la comunicazione end to end;
- Applicazione
  definisce come una applicazione può utilizzare l'insieme dei protocolli TCP/IP.

INTERNET: IL MODELLO TCP/IP

Lo schema mostra la disposizione dei livelli appena descritti:



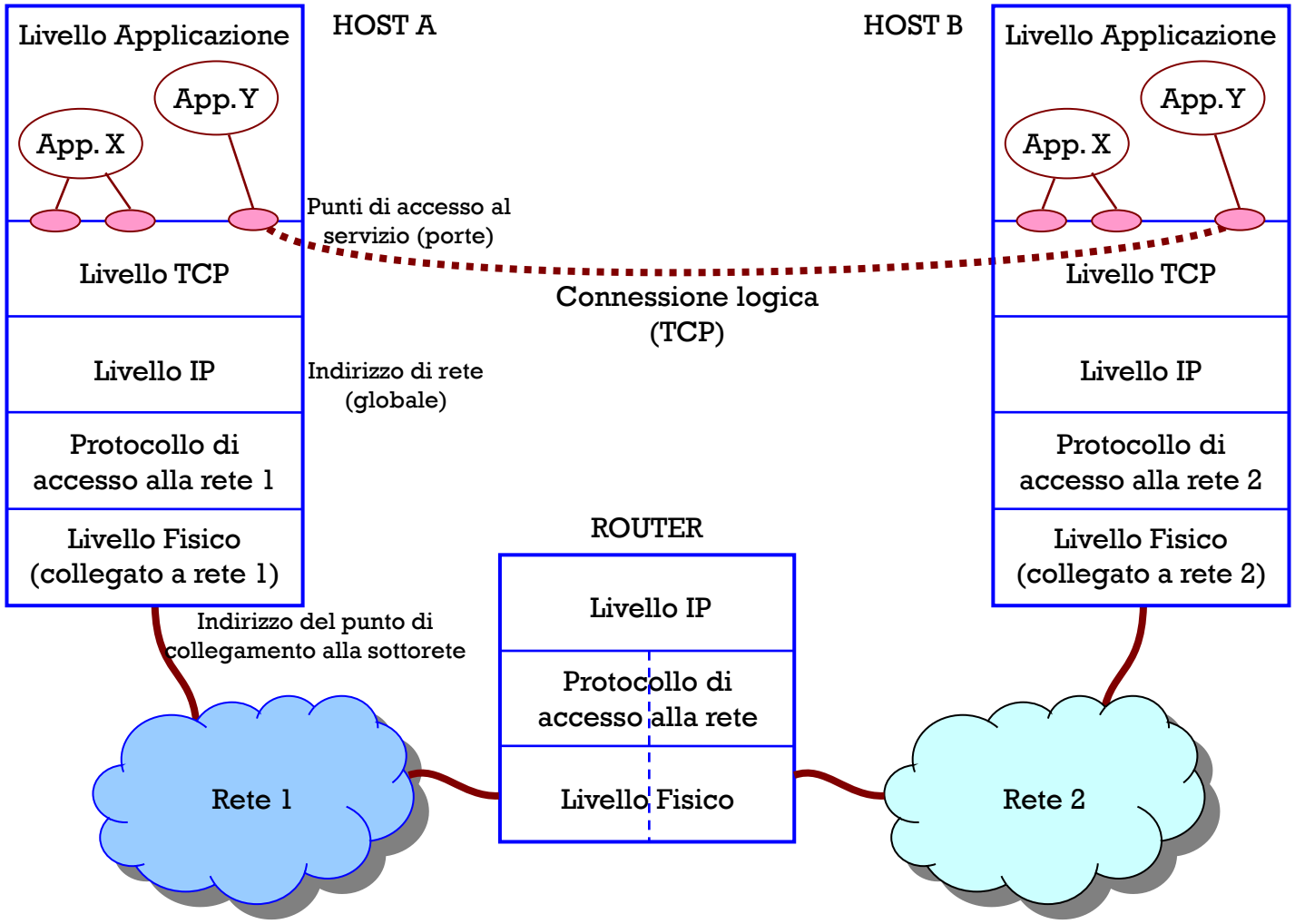
È il primo livello e rappresenta l'interfaccia con l'utente (ad esempio la consultazione di pagine web)

Per gestire più processi nello stesso nodo, fa uso di porte (da 0 a 65535)

Gestisce indirizzamento ed instradamento dei pacchetti

Definisce lo scambio dati tra lo strato fisico e la rete al quale è collegato

INTERNET: IL MODELLO TCP/IP



INDIRIZZAMENTO

Abbiamo visto in precedenza come qualsiasi nodo che intenda inviare dei dati ad un altro nodo, attraverso una rete, deve sempre conoscerne l'indirizzo fisico.

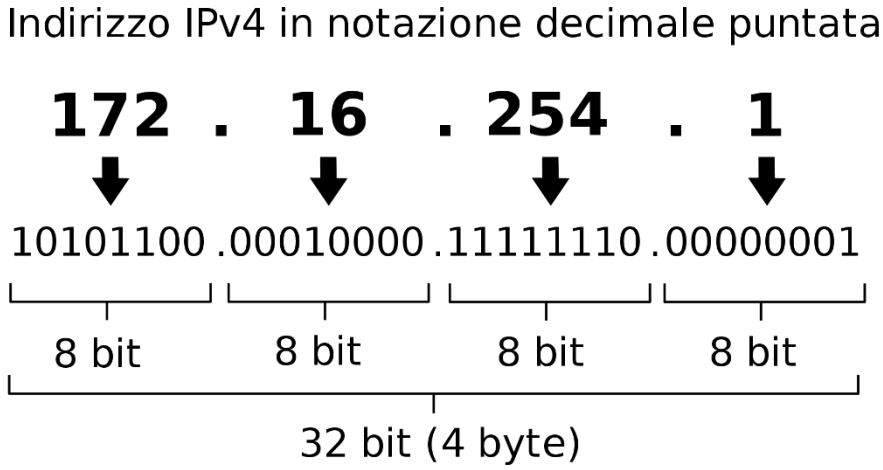
Il meccanismo di indirizzamento, quindi:

- definisce il *formato* degli indirizzi assegnati ai nodi di una rete;
- specifica le *modalità* con le quali gli indirizzi vengono assegnati ai calcolatori (deve garantirne l'univocità).

Reti diverse utilizzano meccanismi di indirizzamento diversi.

INDIRIZZAMENTO IP

Nel caso di internet, come visto, si fa uso degli *indirizzi IP.* Un indirizzo IPv4 è grande 32 bit ed è del tipo:



Dove ogni gruppo di 8 bit (1 byte) è detto *ottetto* e varia, in binario, da 00000000 a 11111111 (da 0 a 255 in decimale)

INDIRIZZAMENTO IP

Un indirizzo IPv6, invece, è grande 128 bit (16 byte) ed è descritto da 8 gruppi di 4 cifre esadecimali (grandi ciascuno 2 byte e quindi compreso tra 0 e 65526), ovvero del tipo:

2001:0DB8:0000:0000:0000:0000:0000:0001

(come si vede si utilizzano i due punti come separatore).

Gli indirizzi possono essere pubblici o privati, inoltre:

- *statici* (ad un nodo viene assegnato sempre lo stesso valore);
- *dinamici* (vengono assegnati dinamicamente da un server della rete tramite il servizio DHCP).

PROTOCOLLO TCP/IP

Come visto, Internet utilizza una suite di protocolli basati sui due più importanti protocolli: il protocollo TCP ed il protocollo IP.

Il TCP è un protocollo a pacchetto che agisce al livello trasporto ed ha il compito di rendere affidabile la comunicazione dei dati (evitare che tra il nodo che invia e quello che riceve, non si perdano pacchetti).

L'IP, invece, è il protocollo che in internet si occupa di indirizzamento e di instradamento dei pacchetti. È nato per connettere tra loro reti eterogene per prestazioni, tecnologia, gestione, etc.

I DOMINI INTERNET

Esistono delle convenzioni ben precise per definire gli indirizzi dei nodi senza ricorrere a valori numerici, ma utilizzando stringhe mnemoniche.

Internet è logicamente organizzata in maniera gerarchica ed è divisa in *domini principali*, uno per ogni nazione. I domini principali, a loro volta, sono a loro volta suddivisi in *domini di primo livello*, *domini di secondo livello* e *domini di terzo livello*.

Il *dominio completo* di un nodo si ottiene navigando la struttura ad albero corrispondente, dalla radice fino alle foglie (i vari domini), separando tra di loro i livelli con il simbolo ".".

I DOMINI INTERNET

Ad esempio, l'indirizzo del Dipartimento di Ingegneria Civile e Architettura dell'Università di Catania è:

dicar.unict.it

Dove:

- **it** è il nome logico che indica il dominio Italia (dominio principale)
- **unict** indica il dominio di primo livello dell'Università di Catania
- **dicar** il dominio di secondo livello associato al DICAR.

In modo analogo gli indirizzi del DIEEI e del Dipartimento di Fisica dell'Università di Catania sono rispettivamente:

dieei.unict.it

dsf.unict.it

I DOMINI INTERNET

I domini principali "nazionali" sono facilmente individuabili in quanto usano le relative sigle:

.it, .fr, .uk, .de, .jp, .es, .ch, …

Poi ci sono i domini US o "internazionali" che invece fanno riferimento al tipo di attività svolta:

.com	(dominio commerciale);
.org	(dominio organizzazioni "no-profit");
.edu	(dominio università americane);
.mil	(dominio militare americano);
.gov	(ente governativo);
.net	(fornitori di servizi di rete).

I DOMINI INTERNET

Altri esempi di indirizzi sono:

- cs.ubc.ca
- informatik.tu-munchen.de
- cs.stanford.edu
- jpl.nasa.gov
- whitehouse.gov
- ibm.com, apple.com, sun.com, hp.com
- un.org
- cern.ch
- murst.it
- lastampa.it
- to.comune.it

DNS

Sono nati per evitare che l'utente debba ricordate a memoria le sequenze di numeri corrispondenti all'indirizzo IP (sequenze di bit), sostituendoli con i corrispondenti domini internet.

Tuttavia, un computer è in grado di comprendere solo bit (e quindi l'indirizzo IP), per cui è stato progettato un meccanismo di codifica, chiamato DNS (Domain Name Server) che consente ad un browser di determinare, dato un indirizzo a domini, l'indirizzo IP corrispondente.

WORLD WIDE WEB

Il WWW è un servizio messo a disposizione attraverso internet, costituito da un enorme ipertesto distribuito a livello mondiale in cui si possono trovare vari tipi di documenti e informazioni.

Il WWW è basato su uno schema client/server.

- I server mettono a disposizione di chiunque acceda alla rete, servizi e informazioni.
- Un client WWW è una qualunque macchina che permette di accedere a tali servizi attraverso un browser.

INDIRIZZI URL

Il browser è il programma che permette di collegarsi al www. I più diffusi, oggi, sono:

- Mozilla Firefox;
- Microsoft Explorer (sostituito da edge);
- Google Chrome.

Mediante il browser si accede alle pagine HTML. Bisogna specificare solamente l'indirizzo della pagina che si vuole visualizzare, in formato URL (Uniform Resource Locator).

INDIRIZZI URL

http://www.unict.it/index.html

Nome del protocollo

Indirizzo (a domini)

Pathname pagina HTML

INDIRIZZI URL

http://www.repubblica.it/index.html

Nome del protocollo	Indirizzo (a domini)	Pathname pagina HTML

http://www.dmi.unict.it/salfi/fondamentiing.htm

Nome del protocollo	Indirizzo (a domini)	Pathname pagina HTML

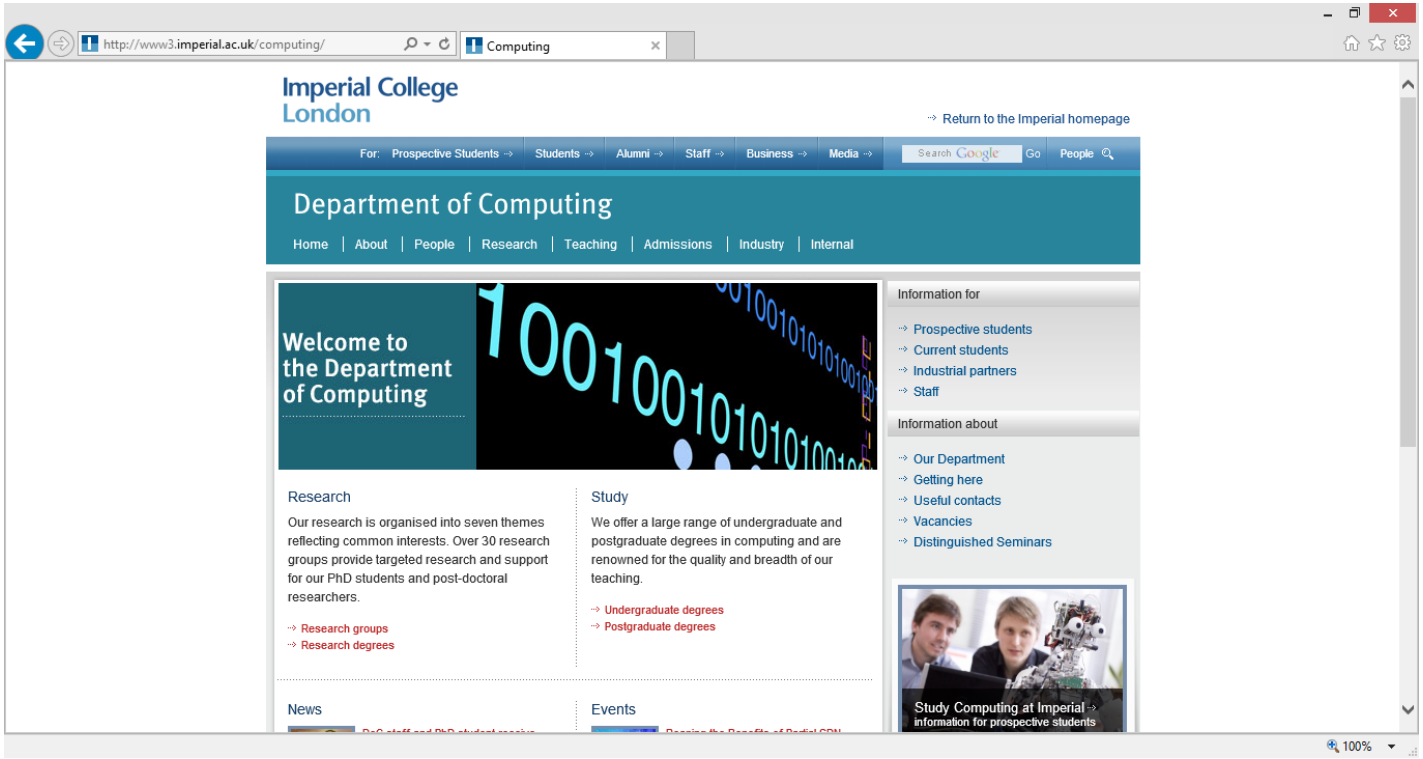
NAVIGARE IN RETE

Che cosa osserviamo quando (mediante il browser) si accede ad una pagina HTML ?

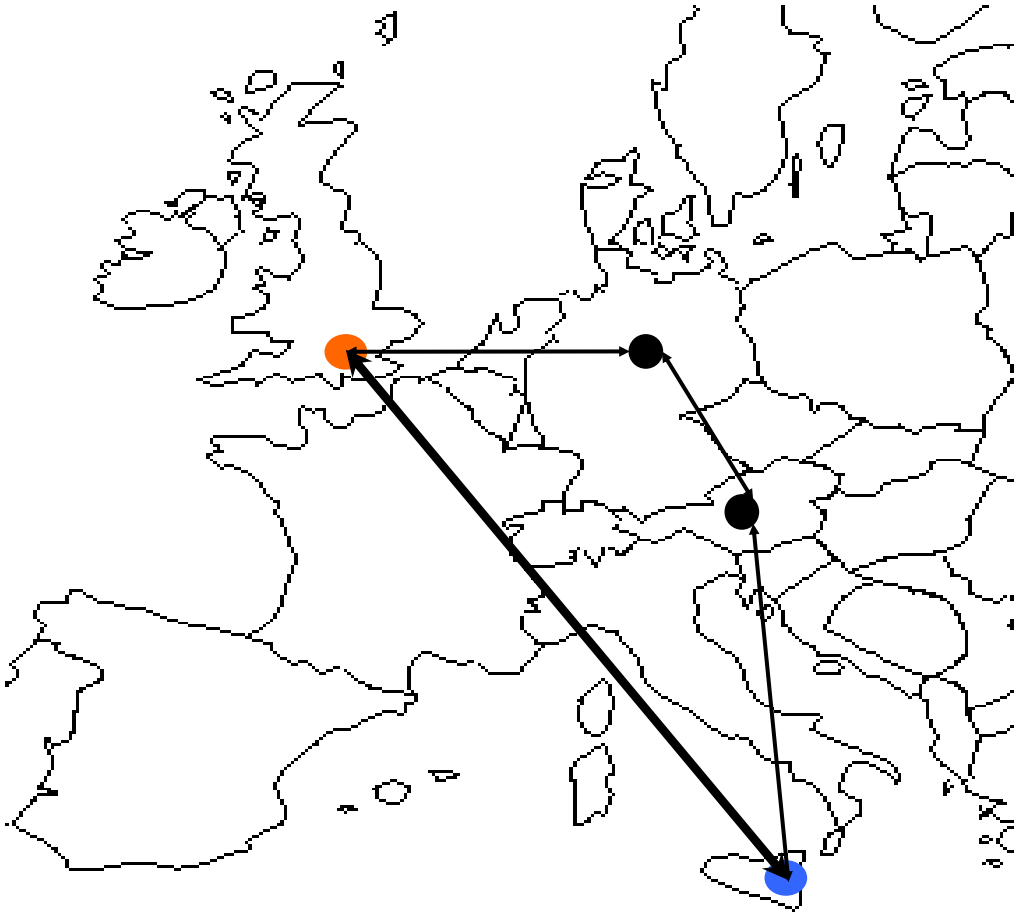
- Si ha l'impressione di avere un collegamento diretto con il computer che viene specificato mediante l'URL.
- Attivando un link (utilizzando il mouse) si accede ad un'altra pagina HTML.

IPERTESTI

I link possono essere sia in forma testuale che grafica:



COSA ACCADE ?



SERVIZI PER GLI UTENTI

Il www non è l'unico servizio offerto dalla rete Internet. Tra gli altri, vi sono i seguenti servizi:

- Posta Elettronica (e-mail).
- Trasferimento File (FTP).
- Connessione remota (Telnet).

POSTA ELETTRONICA

Gli utenti internet possono scambiarsi messaggi utilizzando la posta elettronica (E-mail).

Per inviare un messaggio è sufficiente fornire:

- l'indirizzo del destinatario;
- corpo del messaggio.

Il messaggio, di solito, viene recapitato in modo pressoché istantaneo.

POSTA ELETTRONICA

Gli indirizzi degli utenti vengono stabiliti seguendo le stesse regole per la definizione degli indirizzi dei nodi.

massimiliano.salfi@unict.it

Un indirizzo è formato da due parti separate dal simbolo @ (che si legge «at»):

- nome dell'utente (massimiliano.salfi nell'esempio sopra);
- dominio associato al server (sul quale è installato il software di gestione della posta elettronica), presso cui l'utente è accreditato.

POSTA ELETTRONICA

Per la gestione della posta elettronica da parte di un utente, esistono vari software applicativi:

- Outlook;
- Thunderbird;
- Opera Mail;
- …

È possibile salvare i messaggi ricevuti, visualizzare il contenuto della cassetta della posta, rispondere ad un messaggio, etc. Tipicamente i software di gestione e-mail supportano le funzioni di base:

- Composizione;
- Trasferimento;
- Notifica;
- Eliminazione.

EMAIL: COMPOSIZIONE

Si riferisce al processo di creazione di nuovi messaggi e risposte a messaggi ricevuti.

A prescindere dal software utilizzato, per l'invio di una email è necessario fornire almeno i seguenti campi:

- Mittente;
- Destinatario;
- Oggetto;
- Cc (Copia carbone);
- Ccn (Copia carbone nascosta).

EMAIL: TRASFERIMENTO

Si riferisce al processo di trasferimento dei messaggi dal mittente al destinatario.

Ogni sistema di posta elettronica svolge in maniera automatica la connessione verso il destinatario ed il successivo invio.

EMAIL: NOTIFICA

Di solito è necessario notificare al mittente ciò che è accaduto al messaggio.

- È stato spedito correttamente?
- È andato perso?
- È stato rifiutato dal server ricevente?

Esistono diversi casi per i quali è importante la "ricevuta di ritorno".

EMAIL: ELIMINAZIONE/ARCHIVIAZIONE

È il passo finale nel ciclo di vita di una email e prevede la definizione di quale recipiente usare per il messaggio dopo la ricezione.

Tra le diverse possibilità:

- Spostarlo nel cestino prima di leggerlo;
- Spostarlo nel cestino dopo averlo letto;
- Eliminarlo definitivamente;
- Salvarlo per scopi futuri.

FILE TRANSFER PROTOCOL

Il protocollo FTP permette il trasferimento di files da una macchina ad un'altra (anche su reti geografiche).

È previsto un sistema di protezione dei dati attraverso:

- Autorizzazione di macchine;
- Autorizzazione di utenti (mediante password).

I MOTORI DI RICERCA

Sono dei siti Web che permettono di effettuare ricerche nel World Wide Web attraverso l'indicizzazione dei contenuti esistenti all'interno dei vari siti internet presenti in rete.

In genere tengono conto anche dell'esperienza nelle precedenti ricerche, dei siti più cliccati, ma non è escluso che forniscano anche siti sponsorizzati.

Tra i più diffusi abbiamo:

- www.google.com
- www.altavista.com
- www.excite.com
- www.virgilio.it
- www.bing.it

LA SICUREZZA IN RETE

Con milioni di comuni cittadini che utilizzano le reti per operazioni bancarie, commerciali e fiscali, è necessario garantire la segretezza e l'integrità dei dati.

Rendere sicura una rete, però, non vuol dire solo mantenerla libera da errori di programmazione.

LA SICUREZZA IN RETE

Una rete «sicura», per esempio, è quella in cui nessuno possa leggere e/o modificare i dati destinati ad altri.

La sicurezza implica una lotta contro avversari intelligenti, che sono tecnologicamente ben attrezzati.

LA SICUREZZA IN RETE

I problemi di sicurezza si suddividono in 4 aree:

- Segretezza:

  riservatezza delle informazioni nei confronti degli utenti non autorizzati.

- Autenticazione:

  determinare con chi si sta parlando prima di rivelare informazioni particolari, o iniziare una trattativa d'affari.

- Non Disconoscimento:

  riconoscere le "firme", per essere sicuri che chi ha spedito un messaggio non possa negare di averlo fatto.

- Controllo di Integrità:

  accertare che un messaggio sia davvero quello spedito e non qualcosa di modificato o inventato.

MALWARE

Un malware è un programma informatico realizzato allo scopo di danneggiare il sistema infettato.

Con lo sviluppo di internet la diffusione dei virus è aumentata in maniera esponenziale. Il maggior veicolo di trasporto sono le e-mail.

MALWARE

Un malware può essere allegato ad una e-mail. L'utente ignaro apre l'e-mail per leggerne il contenuto attivando, in modo inconsapevole, il virus stesso.

I danni che possono essere causati dai malware vanno da semplici scritte, più o meno simpatiche, alla perdita completa dei dati, o al blocco del sistema.

MALWARE

- **Virus**

  Si diffondono copiandosi all'interno di altro software. Di solito vanno in esecuzione quando si esegue il programma infetto.

- **Worm**

  Non infettano altri programmi, ma si installano nel SO per cui vanno in esecuzione autonomamente.

- **Trojan horse**

  Software che, oltre ad avere parti lecite che inducono l'utente all'installazione, svolge azioni dannose.

- **Backdoor**

  Consentono un accesso non autorizzato al Sistema (una sorta di porta sul retro).

- **Spyware**

  Raccolgono informazioni dal pc infetto e le inviano ad un computer remoto.

- **Ransomware**

  Limita l'accesso del dispositivo infetto, chiedendo un riscatto (ransom) per la rimozione.

MALWARE

- **Dialer**

  Dirottano la connessione telefonica verso numeri internazionali.

- **Hijacker**

  Si appropriano del browser e dirottano la navigazione verso siti "pirati"

- **Scareware**

  Ingannano l'utente, facendogli credere di avere il computer infetto, invitandolo ad installare del software (di tipo malware).

- **Rabbit**

  Esauriscono rapidamente le risorse del computer infetto, creando rapidamente un numero elevato di copie.

- **Batch**

  Virus "amatoriali" che di solito non arrecano grossi danni.

- **Keylogger**

  Ispirati ai sistemi di log (utilissimi in contesto sistemistico) intercettano per scopi non leciti tutto quello che viene digitato sulla tastiera di un computer infetto.

ANTI-VIRUS

In commercio si trovano diversi applicativi software che agiscono a protezione del sistema, rilevando e ripulendo svariati tipi di malware. È buona abitudine aggiornare spesso l'antivirus, poiché molto frequentemente appaiono nuovi «ceppi virali».

Esempi di antivirus:

- Norton antivirus,
- McAfee antivirus.
- Panda;
- AVG;
- ….

ANTI-VIRUS

Nessun antivirus, tuttavia, dà garanzia assoluta di non subire contagi.

Solitamente, oramai, gli antivirus si limitano ad intercettare comportamenti anomali ed a segnalarli e/o bloccarli.

REGOLE DI CONDOTTA GENERALE

È sempre buona norma:

- non scaricare software arbitrario da Internet;
- non navigare in siti di dubbia origine;
- non aprire allegati se non si è certi sulla provenienza del messaggio, in caso di dubbi, contattare prima il mittente della email;
- evitare di connettere al proprio computer hard disk esterni o pen drive di provenienza dubbia;
- usare password efficaci e cambiarle almeno una volta ogni 15 giorni;

REGOLE DI CONDOTTA GENERALE

Come regola generale per le password:

- evitare di utilizzare le stesse password per ogni account importante (ad esempio login computer, home banking, posta elettronica, social network, etc.);
- ogni password deve essere lunga almeno 8 caratteri, avere sia lettere maiuscole che minuscole, contenere numeri e caratteri speciali quali @, &, %,!, <, >, $, etc;
- non utilizzare informazioni personali o parole comuni;
- tenere le password al sicuro.

REGOLE DI CONDOTTA GENERALE

Oltre quelli affrontati nelle slides precedenti, esistono altri aspetti connessi alla sicurezza informatica.

L'avvento dei social ha sicuramente cambiato il modo di comunicare. Ma siamo sempre certi di sapere con chi stiamo «chattando»?

ALCUNI ASPETTI LEGALI

- Aprire un profilo falso su Facebook, o sui social in genere, non è reato a meno che si assumano i connotati di una persona **realmente esistente**.
- Ciò che è vietato, dunque, è rubare l'identità a una persona realmente esistente: scatta allora il reato di «**furto d'identità**».
- La creazione di un **fake** per infastidire un'altra persona in chat fa scattare una condanna penale per **molestie**.

ALCUNI ASPETTI LEGALI

- Quando il comportamento diventa insistente e ripetuto, anche se per lassi di tempo non particolarmente lunghi, si può incorrere nel reato di **stalking**.
- Un altro reato è quello commesso quando si utilizza il fake per **adescare una minorenne**, facendole credere di essere più piccoli per avere un incontro a sfondo sessuale. In questo caso scatta anche l'**aggravante** dell'utilizzo del social network come «modalità subdola dell'adescamento»

CYBERBULLISMO

Con il termine «Cyberbullismo» si intende:

- qualunque forma di pressione, aggressione, molestia, ricatto, ingiuria, denigrazione, diffamazione, furto d'identità, alterazione, acquisizione illecita, manipolazione, trattamento illecito di dati personali, solitamente in danno di minorenni, realizzata per via telematica;
- nonché la diffusione di contenuti on line, aventi ad oggetto anche uno o più componenti della famiglia del minore, il cui scopo intenzionale e predominante sia quello di isolare un minore o un gruppo di minori, ponendo in atto un serio abuso, un attacco dannoso, o la loro «messa in ridicolo».

CYBERBULLISMO

La legge n.71 del 29 maggio 2017 prevede:

- **SEGNALAZIONE**
  Ciascun minore con più di 14 anni, genitore o chi esercita la responsabilità, può inoltrare al titolare del trattamento o al gestore del sito internet o del social media, un'istanza per l'**oscuramento, la rimozione o il blocco di qualsiasi dato del minore** vittima di cyberbullismo.
- **RIMOZIONE**
  Se entro le 24 ore successive la segnalazione non vi sia stata alcuna comunicazione da parte del responsabile ed entro le 48 ore non abbia provveduto o non sia possibile individuare il gestore del sito internet o del social, l'interessato può fare **istanza al Garante per la protezione dei dati personali**, il quale provvede entro quarantotto ore alla rimozione.

CYBERBULLISMO

La legge n.71 del 29 maggio 2017 prevede:

- **RUOLO DELLA SCUOLA**
  In ogni istituto tra i professori sarà individuato un **referente** per le iniziative contro il bullismo e il cyberbullismo. Il dirigente scolastico che venga a conoscenza di atti di cyberbullismo, deve informare tempestivamente i soggetti che esercitano la responsabilità genitoriale o i tutori dei minori coinvolti e attivare adeguate **azioni di carattere educativo**.

- **AMMONIMENTO**
  Un provvedimento studiato nella logica di educare e responsabilizzare i giovani che, anche solo inconsapevolmente, si rendono attori di comportamenti **penalmente perseguibili**.
  La procedura prevede che fino a quando non sia stata proposta denuncia per diffamazione, minaccia o trattamento illecito di dati, il Questore è chiamato a convocare il minore, colpevole verso altri minori, unitamente a un genitore, ammonendo il medesimo.

CYBERBULLISMO

La legge n.71 del 29 maggio 2017 prevede:

- **EDUCAZIONE CONTINUA NELLE SCUOLE**
  L'uso consapevole della rete entra a far parte dell'**offerta formativa** in ogni ordine di scuola. Il Ministero dell'Istruzione adotta delle linee d'orientamento per la prevenzione e il contrasto del fenomeno. Gli uffici scolastici regionali sono chiamati a promuovere **progetti elaborati nelle scuole**, nonché azioni integrate sul territorio di contrasto del cyberbullismo e **educazione alla legalità.** E' disposto, inoltre, che le istituzioni scolastiche promuovano, nell'ambito della propria autonomia, l'educazione all'**uso consapevole della rete** internet e ai diritti e doveri ad esso connessi.

CYBERBULLISMO

La legge n.71 del 29 maggio 2017 prevede:

- **TAVOLO INTERMINISTERIALE PERMANENTE**
  Il tavolo ha il compito di redigere un **piano di azione integrato** per la prevenzione e il contrasto del cyberbullismo, nonché quello di realizzare un sistema di raccolta dei dati per monitorare l'evoluzione dei fenomeni anche avvalendosi della **collaborazione con la Polizia postale** e le altre forze di polizia. Il tavolo, coordinato dal MIUR, include i Ministeri dell'Interno, Lavoro e Politiche sociali, Giustizia, Sviluppo Economico e della Salute, Anci, Garante per l'Infanzia e l'Adolescenza, Garante Privacy, Comitato di applicazione del codice di autoregolamentazione media e minori, organizzazioni già coinvolte nel programma nazionale del Safer Internet Centre, nonché una rappresentanza delle associazioni studentesche e dei genitori.

- **RISORSE POLIZIA POSTALE**
  Nell'ambito di ciascun programma operativo nazionale sono stanziate idonee risorse alla formazione del personale specializzato alla tutela dei minori sul web. I fondi certi per la Polizia Postale sono per l'aggiornamento ai docenti, nella chiave di individuare referente cyberbullismo per ogni autonomia scolastica e dare luogo alla formazione continua dedicata agli studenti.

SEXTING

Inizialmente, con il termine «Sexting» si intendeva l'invio di messaggi testuali a contenuto sessualmente esplicito.

Il termine sexting, infatti, deriva dalla fusione delle parole inglesi **sex** (sesso) e **texting** (invio di messaggio, in origine testo elettronico).

L'avvento degli smartphone ha esteso tale pratica alla **condivisione di propri video o immagini di nudo** o sessualmente esplicite, ripresi ed inviati utilizzando lo smartphone, con persone delle quali ci si fida.

SEXTING

Quali sono gli aspetti caratteristici del fenomeno?

- **FIDUCIA**
  spesso i ragazzi/le ragazze inviano proprie immagini o video nudi o sessualmente espliciti, perché **si fidano della persona** a cui stanno inviando tale materiale. Mostrano una scarsa consapevolezza che quello stesso materiale, se il rapporto (di amicizia o di coppia) dovesse deteriorarsi o rompersi, potrebbe essere diffuso come ripicca per quanto accaduto.

- **PERVASIVITA'**
  le possibilità che offrono i telefonini di nuova generazione permettono di condividere le foto proprie o altrui, con molte persone contemporaneamente, attraverso invii multipli, condivisione sui social network, diffusione online.

SEXTING

Quali sono gli aspetti caratteristici del fenomeno?

- **REPLICABILITA'**
  Non si ha a che fare con un documento originale, distrutto il quale si è eliminata ogni traccia. Qualsiasi documento del genere può essere copiato più volte, anche attraverso screenshot, e quindi nuovamente rimesso in rete.

- **RICERCABILITA'**
  Ovvero il grado di reperibilità di una risorsa che finisce su Internet. I documenti presenti in rete **possono essere facilmente reperiti** attraverso l'indicizzazione da parte dei motori di ricerca. Quando il file viene immesso in rete, dunque, ne perdiamo il controllo

SEXTING

Quali sono gli aspetti caratteristici del fenomeno?

- **PERSISTENZA DEL FENOMENO**
  Il materiale pubblicato su internet può rimanere disponibile online anche **per molto tempo**. I ragazzi, che crescono immersi nelle nuove tecnologie, non sono consapevoli che una foto o un video diffusi in rete **potrebbero non essere tolti mai più**.

- **NON CONSAPEVOLEZZA**
  I ragazzi spesso non sono consapevoli di scambiare **materiale pedopornografico**. L'invio di foto che ritraggono minorenni al di sotto dei 18 anni in pose sessualmente esplicite configura, infatti, il reato di distribuzione di materiale pedopornografico.

SEXTING

Alcuni suggerimenti:

- **Parlare** con i propri genitori o con un adulto fidato;
- Inoltrare **denuncia** alla Polizia Postale;
- Contattare una delle tante associazioni che si occupano del fenomeno;
- È possibile chiedere a Google la **rimozione di contenuti indicizzati** nella Ricerca, attraverso il link:

https://www.google.com/webmasters/tools/legal-removal-request?complaint_type=rtbf&visit_id=1-636573110212150434-1533066808&rd=1&pli=1

REVENGE PORN

È un fenomeno strettamente connesso al sexting e consiste in una sorta di **«vendetta porno»**, di solito messa in atto da un ex partner a storia finita, attraverso la condivisione pubblica in rete di immagini intime, senza il preventivo consenso della persona ripresa.

In alcuni casi le immagini, o video, vengono ripresi con il consenso della vittima, in altri a sua insaputa.

Al fine di contrastare il fenomeno è stata introdotta la legge 19 luglio 2019 n. 69 che, all'articolo 10, ha approvato il **«Codice Rosso»**, introducendo l'art.612-ter nel codice penale grazie al quale tale fenomeno è diventato reato (il cosiddetto reato di revenge porn).

REVENGE PORN

È prevista la reclusione da uno a sei anni ed una multa da 5.000 a 15.000 euro per chiunque invii, consegni, ceda, pubblichi o diffonda tali contenuti senza il consenso della persona ritratta.

Tale pena viene inasprita nei seguenti casi:

- l'autore del reato è un partner o un ex;
- i contenuti vengono diffusi tramite strumenti telematici o informatici;
- se la vittima è in condizioni di inferiorità fisica o psichica;
- se la vittima è una donna in gravidanza.

INTRODUZIONE ALLE BASI DI DATI

SISTEMA INFORMATIVO

È un componente (un sottosistema) presente in quasi tutte le organizzazioni:

- Aziende private;
- Pubblica amministrazione;
- etc.

Gestisce (acquisisce, elabora, conserva, produce) le informazioni di interesse (ad esempio utilizzate per il perseguimento degli scopi dell'organizzazione stessa):

- ogni organizzazione ha un sistema informativo, eventualmente non esplicitato nella struttura;
- quasi sempre il sistema informativo è di supporto ad altri sottosistemi;
- il sistema informativo è di solito suddiviso in sottosistemi (in modo gerarchico o decentrato), più o meno fortemente integrati.

SISTEMA ORGANIZZATIVO

È l'insieme delle risorse e delle regole messe in atto per lo svolgimento coordinato delle attività, al fine del perseguimento degli scopi:

- il sistema informativo è parte del sistema organizzativo;
- il sistema informativo esegue/gestisce processi informativi (cioè i processi che coinvolgono informazioni).

RISORSE AZIENDALI

Le risorse di una azienda, in genere, sono:

- persone (dette anche risorse umane);
- denaro;
- materiali (o comunque beni);
- **informazioni.**

Oggi viviamo nell'era della conoscenza. Il capitale principale delle organizzazioni è rappresentato dalla conoscenza (basata sui dati).

SISTEMI INFORMATIVI E AUTOMAZIONE

Il concetto di "sistema informativo" è indipendente da qualsiasi automazione.

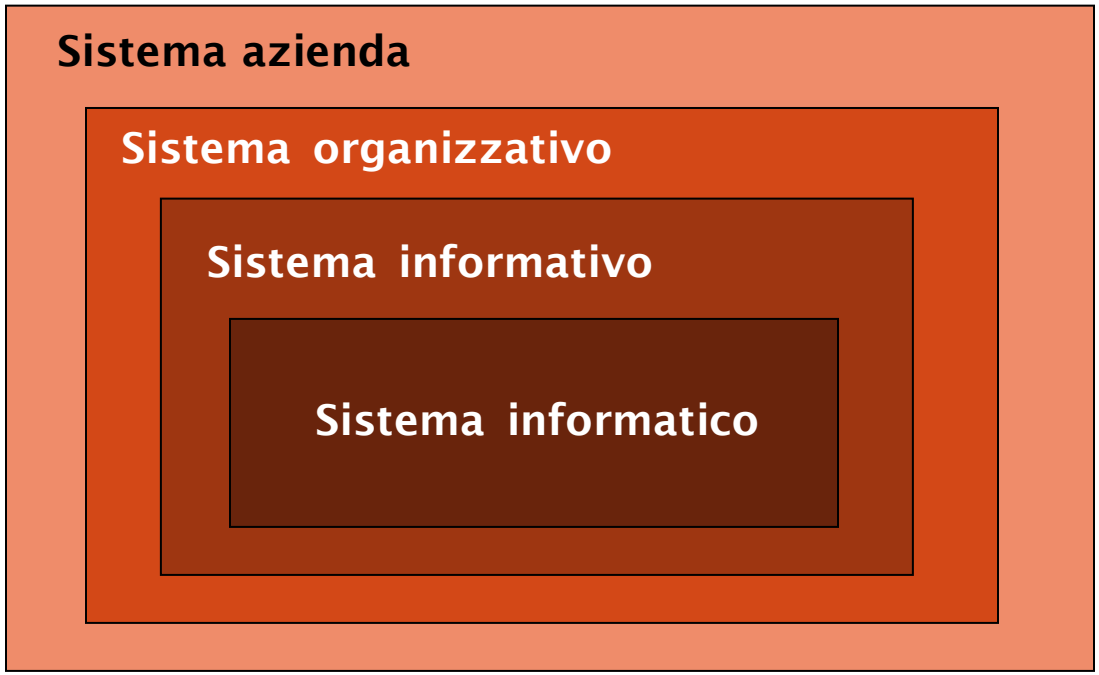
Esistono organizzazioni la cui ragion d'essere è la gestione di informazioni. Ad esempio:

- i servizi anagrafici;
- i tribunali;
- le banche;
- etc.

Esse operano da secoli, pertanto la gestione delle loro informazioni si è evoluta nel tempo senza che sia stato alterato in alcun modo l'obiettivo dell'organizzazione.

SISTEMA INFORMATICO

Rappresenta la porzione automatizzata del sistema informativo, vale a dire la parte del sistema informativo che gestisce le informazioni attraverso la tecnologia informatica.



SISTEMA INFORMATIVO E SISTEMA INFORMATICO

Anche prima di essere informatizzati, molti sistemi informativi si sono evoluti verso una razionalizzazione ed una standardizzazione delle procedure e dell'organizzazione delle informazioni.

Esempio: uffici anagrafe di Catania, passati dalla gestione cartacea a quella elettronica degli atti.

GESTIONE DELLE INFORMAZIONI

Nelle attività umane, le informazioni vengono gestite (registrate e scambiate) in forme diverse:

- idee informali;
- linguaggio naturale (scritto o parlato, formale o colloquiale, in una qualsiasi lingua);
- disegni, grafici, schemi;
- numeri e codici;

e su vari supporti:

- memoria umana;
- carta;
- dispositivi elettronici.

GESTIONE DELLE INFORMAZIONI

Nelle attività standardizzate dei sistemi informativi complessi, sono state introdotte col tempo forme di organizzazione e codifica delle informazioni.

Ad esempio, nei servizi anagrafici si è iniziato con registrazioni discorsive, poi, nel tempo:

- nome e cognome;
- estremi anagrafici;
- codice fiscale.

INFORMAZIONI E DATI

Nei sistemi informatici (e non solo), le informazioni vengono rappresentate in modo essenziale, spartano, attraverso i **dati**.

Dal Vocabolario della lingua italiana (1987):

informazione: notizia, dato, o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere.

dato: ciò che è immediatamente presente alla conoscenza, prima di ogni elaborazione; (in informatica) elementi di informazione costituiti da simboli che debbono essere elaborati.

L'informazione può essere vista come un'interpretazione semantica dei dati che dipende dal contesto.

DATI E INFORMAZIONI

I dati hanno bisogno di essere interpretati. Ad esempio:

`'Mario' '275'` su un foglio di carta sono due **dati.**

Se il foglio di carta viene fornito in risposta alla domanda "A chi mi devo rivolgere per il problema X? Qual è il suo interno?", allora i dati possono essere interpretati per fornire **informazione** e arricchire la **conoscenza.**

Ma il significato può essere differente, ad esempio:

- quanti soldi mi deve Mario?
- quanti giorni ha lavorato Mario?
- etc.

PERCHÉ I DATI?

La rappresentazione precisa di forme più ricche di informazione e conoscenza è piuttosto difficile.

I dati costituiscono spesso una risorsa strategica, perché più stabili nel tempo di altre componenti (processi, tecnologie, ruoli umani, etc.)

SISTEMA DI GESTIONE DI BASI DI DATI
DATABASE MANAGEMENT SYSTEM — DBMS

È un sistema (ovvero un prodotto **software** residente su di un supporto **hardware**) in grado di gestire **collezioni di dati** che siano (anche):

- **Grandi**, di dimensioni (molto) maggiori della memoria centrale dei sistemi di calcolo utilizzati;
- **Persistenti**, con un periodo di vita indipendente dalle singole esecuzioni dei programmi che le utilizzano;
- **Condivise,** utilizzate da applicazioni diverse, anche in località geografiche diverse;
- **Affidabili**, resistenti a malfunzionamenti hardware e software, blackout, etc.
- **Private**, con una disciplina e un controllo degli accessi;

Come ogni prodotto informatico, un DBMS deve essere:

- **Efficiente**, utilizzando al meglio le risorse di spazio e tempo del sistema
- **Efficace**, rendendo produttive le attività dei suoi utilizzatori

ALCUNI DBMS IN COMMERCIO

Quello del software DBMS è un business estremamente grande. Di seguito alcuni tra i tanti sistemi esistenti:

- Microsoft Access;
- IBM-DB2;
- Oracle;
- Informix;
- Sybase;
- Microsoft SQLServer;
- Ingres;
- MySql (open-source).

CONDIVISIONE

Ogni organizzazione (soprattutto se molto grande) è divisa in settori, o comunque svolge diverse attività.

A ciascun settore o attività corrisponde un (sotto)sistema informativo.

Possono esistere sovrapposizioni fra i dati di interesse dei vari settori.

Una base di dati è una risorsa **integrata**, condivisa fra i vari settori.

POSSIBILI PROBLEMI

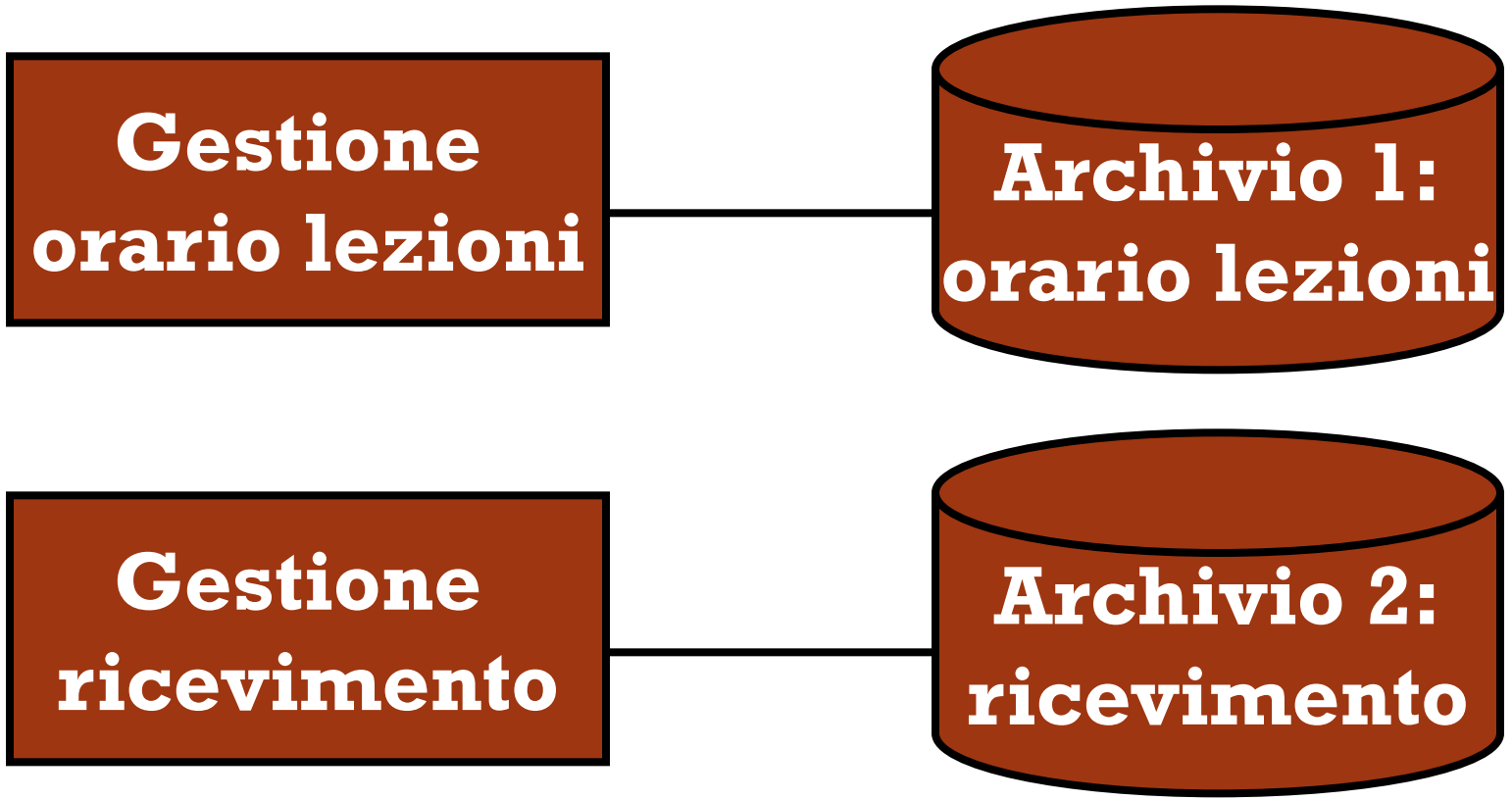
Ridondanza:

- Le informazioni possono essere ripetute più volte all'interno di un database.

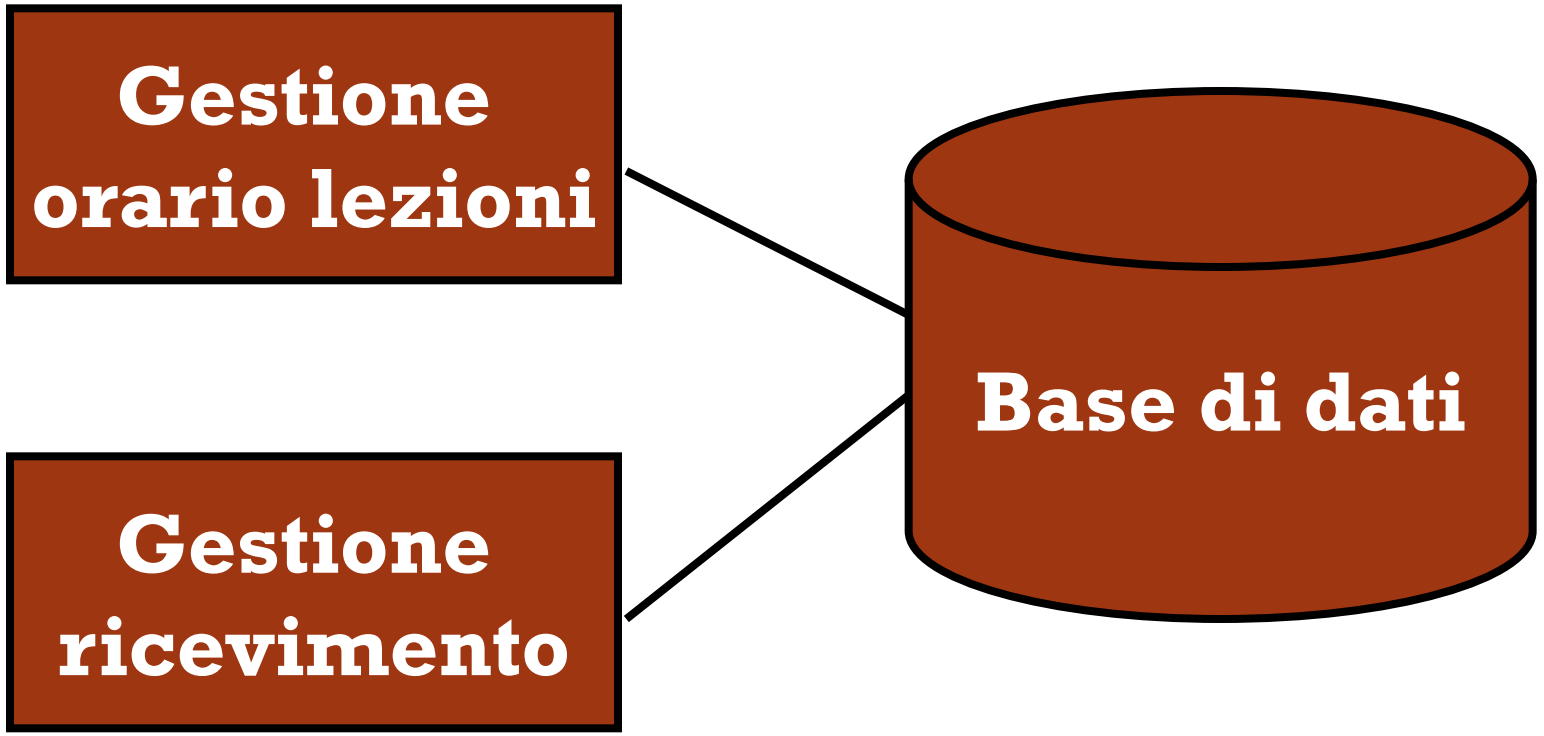
Rischio di **incoerenza**:

- le versioni possono non coincidere.

ARCHIVI E BASI DI DATI



ARCHIVI E BASI DI DATI



LE BASI DI DATI SONO CONDIVISE

Una base di dati e' una risorsa **integrata**, **condivisa** fra le varie applicazioni.

Conseguenze:

- Attività diverse su dati in parte condivisi:
  - meccanismi di **autorizzazione.**
- Attività multi-utente su dati condivisi:
  - controllo della **concorrenza**

EFFICIENZA

Si misura (come in tutti i sistemi informatici) in termini di tempo di esecuzione (tempo di risposta) e spazio di memoria (sia principale che secondaria).

I DBMS, a causa della varietà di funzioni, non sono necessariamente più efficienti dei file system.

L'efficienza è il risultato della qualità del DBMS e delle applicazioni che lo utilizzano.

Organizzazione dei dati in una base di dati

Orario

Insegnamento	Docente	Aula	Ora
Analisi matem. I	Luigi Neri	N1	8:00
Basi di dati	Piero Rossi	N2	9:45
Chimica	Nicola Mori	N1	9:45
Fisica I	Mario Bruni	N1	11:45
Fisica II	Mario Bruni	N3	9:45
Sistemi inform.	Piero Rossi	N3	8:00

Lo **schema** della base di dati

Insegnamento	Docente	Aula	Ora
Analisi matem. I	Luigi Neri	N1	8:00
Basi di dati	Piero Rossi	N2	9:45
Chimica	Nicola Mori	N1	9:45
Fisica I	Mario Bruni	N1	11:45
Fisica II	Mario Bruni	N3	9:45
Sistemi inform.	Piero Rossi	N3	8:00

L'**istanza** della base di dati

SCHEMI E ISTANZE

In ogni base di dati esistono:

- lo **schema**, sostanzialmente invariante nel tempo, che ne descrive la struttura (l'aspetto intensionale).
  Nell'esempio, le intestazioni delle tabelle.

- l'**istanza**, cioè i valori attuali, che possono cambiare anche molto rapidamente (aspetto estensionale).
  Nell'esempio, il "corpo" di ciascuna tabella

campo

Sistemi inform.	Piero Rossi	N3	8:00

Record

UNA VISTA

Corsi

Corso	Docente	Aula
Basi di dati	Rossi	DS3
Sistemi	Neri	N3
Reti	Bruni	N3
Controlli	Bruni	G

Aule

Nome	Edificio	Piano
DS1	OMI	Terra
N3	OMI	Terra
G	Pincherle	Primo

CorsiSedi

Corso	Aula	Edificio	Piano
Sistemi	N3	OMI	Terra
Reti	N3	OMI	Terra
Controlli	G	Pincherle	Primo

LINGUAGGI PER BASI DI DATI

Un altro contributo all'efficacia è dovuto alla disponibilità di vari linguaggi e di interfacce diverse.

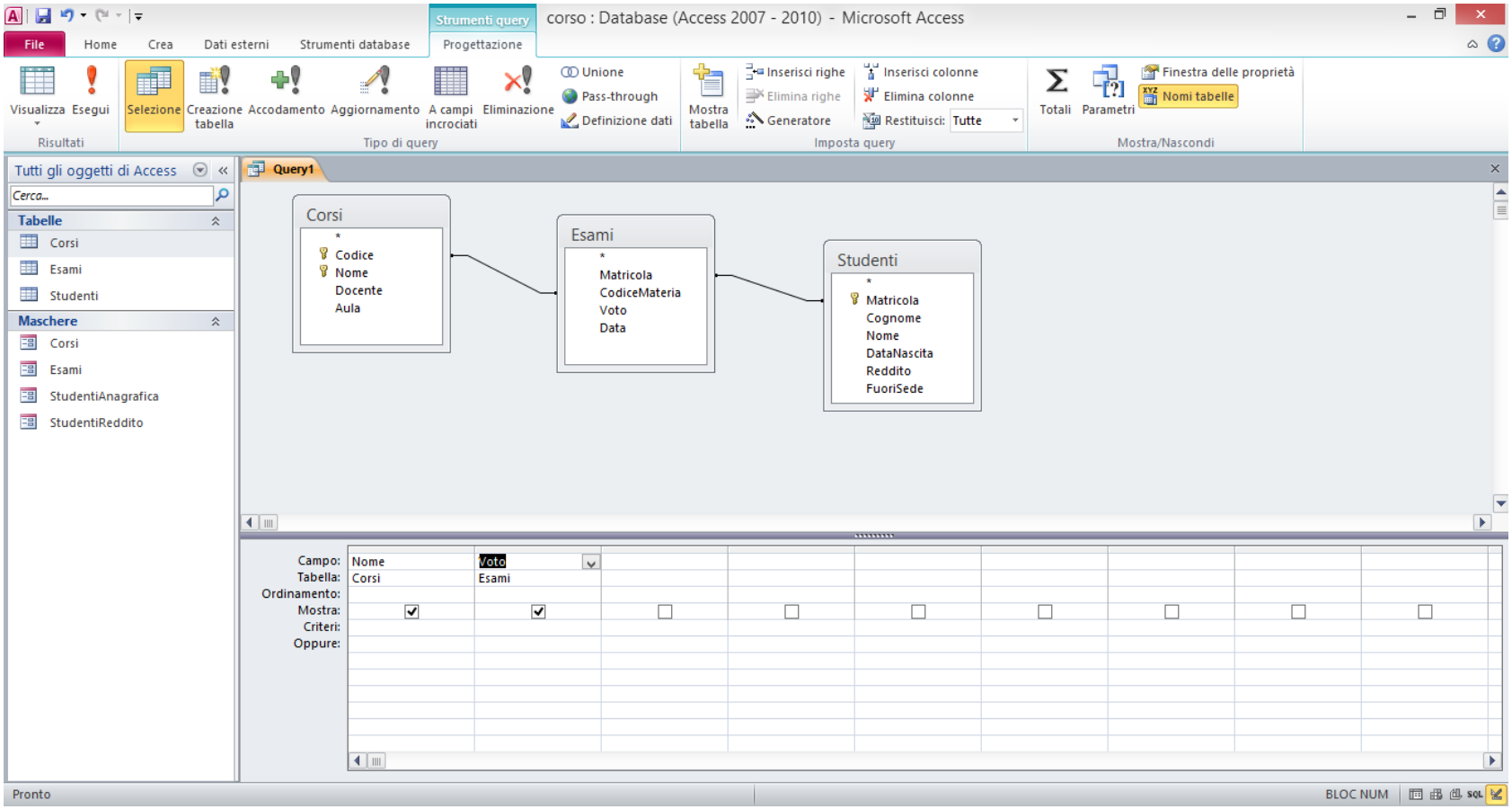
- linguaggi testuali interattivi (**SQL**);
- comandi (come quelli del linguaggio interattivo) immersi in un linguaggio **ospite** (Pascal, C, Cobol, etc.);
- comandi (come quelli del linguaggio interattivo) immersi in un linguaggio ad hoc, con anche altre funzionalità (p.es. per grafici o stampe strutturate), anche con l'ausilio di strumenti di sviluppo (p. es. per la gestione di maschere);
- con interfacce amichevoli (senza linguaggio testuale).

SQL, UN LINGUAGGIO INTERATTIVO

```
SELECT Corso, Aula, Piano
FROM Aule, Corsi
WHERE Nome = Aula
AND Piano="Terra"
```

Corso	Aula	Piano
Sistemi	N3	Terra
Reti	N3	Terra

INTERAZIONE NON TESTUALE (IN ACCESS)



UNA DISTINZIONE TERMINOLOGICA (SEPARAZIONE FRA DATI E PROGRAMMI)

data manipulation language (DML)

per l'interrogazione e l'aggiornamento di (**istanze** di) basi di dati

data definition language (DDL)

per la definizione di **schemi** (logici, esterni, fisici) e altre operazioni generali

UN'OPERAZIONE DDL (SULLO SCHEMA)

```
CREATE TABLE orario (
        insegnamento        CHAR(20),
        docente             CHAR(20),
        aula                CHAR(4),
        ora                 CHAR(5)
)
```

OPERAZIONI DDL (SULLO SCHEMA)

Aggiungi la colonna Piano, di tipo intero:

```
ALTER TABLE orario
ADD Piano INT
```

Elimina la colonna insegnamento dalla tabella orario:

```
ALTER TABLE orario
DROP COLUMN insegnamento
```

Elimina la tabella orario:

```
DROP TABLE orario
```

TRANSAZIONI

Programmi che realizzano attività frequenti e predefinite, con poche eccezioni, previste a priori.

Esempi:

- versamento presso uno sportello bancario;
- emissione di un certificato anagrafico;
- dichiarazione presso l'ufficio di stato civile;
- prenotazione aerea.

Le transazioni sono di solito realizzate con programmi in linguaggio ospite (tradizionale o ad hoc).

N.B. il termine **transazione** ha un'altra accezione, più specifica: sequenza indivisibile di operazioni (le quali devono essere eseguite tutte insieme, altrimenti annullate in blocco).

VANTAGGI E SVANTAGGI DEI DBMS

Pro

- dati come risorsa comune, base di dati come modello della realtà;
- gestione centralizzata con possibilità di standardizzazione ed "economia di scala";
- disponibilità di servizi integrati;
- riduzione di ridondanze e inconsistenze;
- indipendenza dei dati (favorisce lo sviluppo e la manutenzione delle applicazioni).

Contro

- costo dei prodotti e della transizione verso di essi;
- non scorporabilità delle funzionalità (con riduzione di efficienza).

ALCUNI COSTRUTTI DEL LINGUAGGIO SQL

Per selezionare record da una tabella, la sintassi da utilizzare è la seguente:

SELECT Campo1, Campo2,… Campo n

FROM NomeTabella

Se si desidera che i dati siano filtrati da determinate condizioni, occorre aggiungere:

WHERE Campo1 = Valore 1 AND Campo2 = Valore n, …

Si noti che al posto dell'operatore =, si può utilizzare anche < (minore) o > (maggiore) mentre al posto dell'operatore AND, si può utilizzare OR.

ALCUNI COSTRUTTI DEL LINGUAGGIO SQL

Per eliminare record da una tabella, la sintassi da utilizzare è la seguente:

```
DELETE FROM NomeTabella
```

In questo modo la tabella verrà svuotata. Se si desidera, invece, che la cancellazione avvenga solo sui record che soddisfano determinate condizioni, occorre aggiungere:

```
WHERE Campo1 = Valore 1 AND Campo2 = Valore 2 AND …
```

Si noti che al posto dell'operatore = si può utilizzare anche < (minore) o > (maggiore), mentre al posto dell'operatore AND si può utilizzare OR.

ALCUNI COSTRUTTI DEL LINGUAGGIO SQL

Per modificare campi da uno o più record di una tabella, la sintassi da utilizzare è la seguente:

UPDATE NomeTabella

SET Campo1 = Valore1, Campo2 = Valore 2,…

In questo modo la modifica avverrà su tutti i record. Se si desidera, invece, che avvenga solo sui record che soddisfano determinate condizioni, occorre aggiungere:

WHERE Campo 1 = Valore 1 AND Campo2 = Valore2 AND …

Si noti che al posto dell'operatore = si può utilizzare anche < (minore) o > (maggiore), mentre al posto dell'operatore AND si può utilizzare OR.

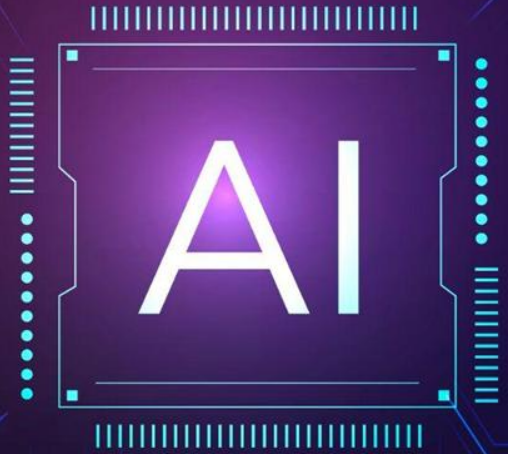
ALCUNI COSTRUTTI DEL LINGUAGGIO SQL

Per inserire record in una tabella, la sintassi da utilizzare è la seguente:

```
INSERT INTO NomeTabella
VALUES (ValoreCampo1, ValoreCampo2, …, ValoreCampo n)
```

In questo caso si dovranno specificare tutti i valori dei campi previsti dallo schema, nell'ordine esatto. Se si desidera, invece, indicare quali campi inserire, la sintassi sarà:

```
INSERT INTO NomeTabella
(Campo1, Campo2,…)
VALUES (ValoreCampo1, ValoreCampo2,…)
```



Elementi di Intelligenza Artificiale ed esempi di utilizzo dei più diffusi chatbot

Andrea Caruso

Andrea Caruso

OUTLINE

Cos’è l’AI

Un po’ di storia

Campi dell’AI

Applicazioni dell’AI

Tipi di apprendimento

Machine Learning

Deep Learning

Reti neurali

Esempi

Cos'è l'AI?

Cos’è l’AI?

Andrea Caruso

Definizione

La norma ISO/IEC 42001:2023 Information technology - Artificial Intelligence Management System (AIMS) definisce l'intelligenza artificiale come la capacità di un sistema (informatico) di mostrare capacità umane quali il ragionamento, l'apprendimento, la pianificazione e la creatività.

Cos’è l’AI?

Quindi l’Intelligenza Artificiale simula, appunto, comportamenti intelligenti…

Ma cos’è l’intelligenza?

Complesso di facoltà psichiche e mentali che consentono di pensare, comprendere o spiegare i fatti o le azioni, elaborare modelli astratti della realtà, intendere e farsi intendere dagli altri, giudicare, e adattarsi all’ambiente.

ENCICLOPEDIA TRECCANI

Cos’è l’AI?

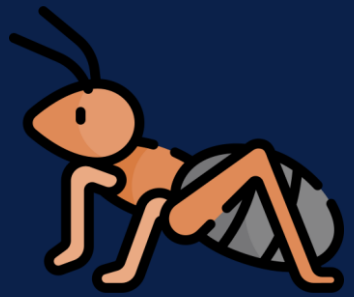
A livello base del comportamento animale, possiamo distinguere tre punti fermi:

- Istinto
- Apprendimento
- Ragionamento

Cos’è l’AI?

Istinto

Le formiche sono un esempio classico di animali guidati principalmente dall'istinto. All'interno di un formicaio, le formiche mostrano un comportamento altamente organizzato e coordinato, come la costruzione di tunnel e camere, la raccolta di cibo e la difesa del formicaio. Questi comportamenti sono innati e guidati principalmente da segnali chimici e meccanismi genetici.



Cos’è l’AI?

Andrea Caruso

Apprendimento

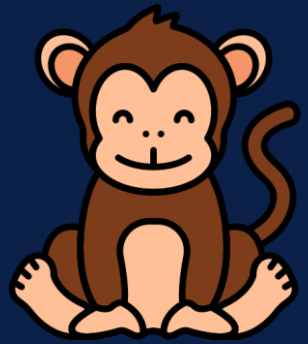
I cani sono noti per la loro capacità di apprendimento attraverso l'addestramento e l'interazione con gli esseri umani. Possono imparare una vasta gamma di comandi e comportamenti, oltre a mostrare una comprensione emotiva delle espressioni umane e delle situazioni.
L'addestramento dei cani si basa su ricompense e punizioni per rinforzare i comportamenti desiderati.

Cos’è l’AI?

Andrea Caruso

Ragionamento

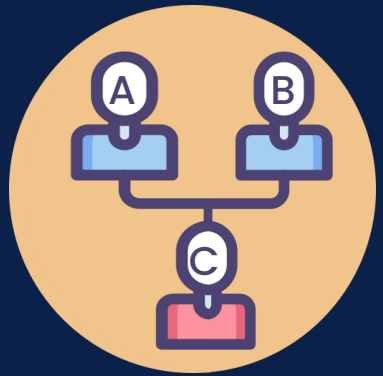
Le scimmie, in particolare gli scimpanzé, sono noti per la loro capacità di ragionamento complesso. Possono risolvere problemi, utilizzare strumenti e pianificare azioni future. Gli scimpanzé dimostrano una forma di ragionamento che va oltre il semplice addestramento, mostrando una capacità di comprendere relazioni causali e di pianificare attività complesse per raggiungere determinati obiettivi.



Cos’è l’AI?

Ma come fare a capire se un programma è intelligente?

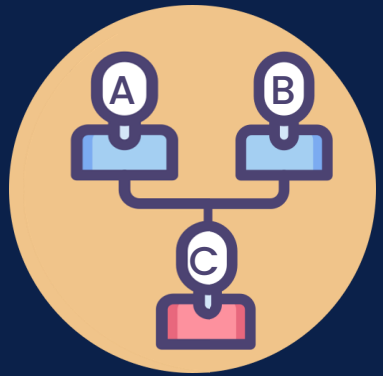
Il Test di Turing è un criterio proposto da A.M.Turing per valutare se una macchina può essere considerata in grado di pensare come un essere umano. Proposto nel 1950 in un articolo pubblicato sulla rivista "Mind", il test consiste di due fasi. La prima fase è in una simulazione di gioco a tre partecipanti: un uomo (A), una donna (B) e una terza persona (C). La terza persona è separata dagli altri due e, tramite una serie di domande, deve determinare chi sia l'uomo e chi la donna (egli sa solo che vi sono un uomo e una donna). Per assicurare che C non abbia indizi visivi o uditivi, le risposte alle domande devono essere trasmesse tramite telescrivente o comunque in modo non riconoscibile.
Le domande possono essere, ad esempio:
“hai i capelli corti?”, “porti la gonna?”, “hai il rossetto?”



Andrea Caruso

Cos'è l'AI?

A e B hanno compiti specifici: A deve ingannare C e indurlo a fare un'identificazione sbagliata, mentre B deve aiutarlo. L'intervistatore non ha idea di chi menta e di chi sia sincero. Deve riuscire a capirlo da sé. Alla fine del gioco l'intervistatore deve decidere chi dei due partecipanti è l'uomo e chi la donna.



Ripetendo il gioco N volte, l'intervistatore sbaglia il sesso dei partecipanti per X volte. Il suo tasso di errore è pari a X/N. Ad esempio, se su 100 partite l'intervistatore sbaglia 30 volte il sesso dei due partecipanti, il suo tasso di errore X/N è pari a 0,30, ossia al 30%.

Andrea Caruso

Cos’è l’AI?

Il Test di Turing si basa sull'idea che una macchina possa sostituire A, e se C non riesce a distinguere la macchina dall'uomo, la macchina è considerata intelligente. Infatti, nella seconda fase del test uno dei due partecipanti viene sostituito da un calcolatore.
Il compito di C è sempre lo stesso: distinguere chi sia l’uomo e chi la donna.
Ripetendo il gioco N volte, l'intervistatore sbaglia il sesso dei partecipanti per Y volte. Il suo tasso di errore è pari a Y/N.

Cos'è l'AI?

Quando il Test di Turing viene considerato superato?

Se, nell'individuare il sesso di A e B, il tasso di errore è simile prima e dopo la sostituzione di A con la macchina, allora la macchina stessa dovrebbe essere considerata intelligente, dal momento che, in questa specifica situazione, è indistinguibile da un essere umano.

In termini matematici, il test è superato quando

$$X/N \simeq Y/N$$

Cos’è l’AI?

Test di Turing semplificato

Ci sono molte varianti del Test di Turing.
In una delle versioni più basilari, un intervistatore umano interagisce tramite un terminale con un'altra entità, la quale potrebbe essere sia una persona, sia una macchina, situata in una stanza separata.
La sfida per la macchina è quella di convincere l'intervistatore che anche lei sia un essere umano durante la conversazione. Se l’uomo non è in grado di distinguere se stia parlando con una macchina o con un intervistatore umano, allora il test è superato.

Cos’è l’AI?

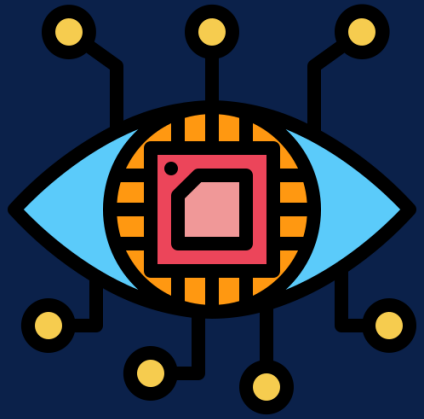
L’opinione di Turing

«Secondo la forma più estrema di questa opinione, il solo modo per cui si potrebbe essere sicuri che una macchina pensa è quello di essere la macchina stessa e sentire se si stesse pensando. [...] Allo stesso modo, la sola via per sapere che un uomo pensa è quello di essere quell'uomo in particolare. [...] Probabilmente A crederà "A pensa, mentre B no", mentre per B è l'esatto opposto "B pensa, ma A no". Invece di discutere in continuazione su questo punto, è normale attenersi alla educata convenzione che ognuno pensi.»

Cos’è l’AI?

Altre varianti del test

Test di Turing totale (2010): Valuta non solo la capacità di risposta scritta della macchina, ma anche quella visiva (visione artificiale) e quella di manipolazione degli oggetti (robotica).



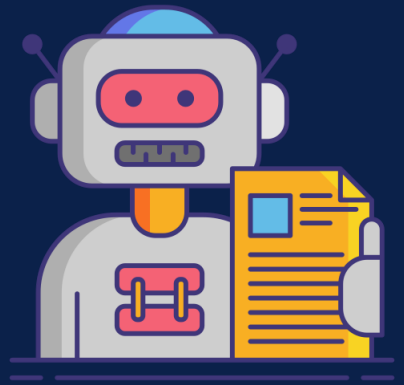
Russell, Stuart J.; Norvig, Peter (2010), Artificial Intelligence: A Modern Approach (3rd ed.), Upper Saddle River, NJ: Prentice Hall

Cos’è l’AI?

Altre varianti del test

Test di Turing dell’esperto (2003): Per superare il test, la macchina deve rispondere in modo da non essere distinguibile da un essere umano esperto nello specifico campo di interesse. L'obiettivo della macchina è quello di indurre l'osservatore a fare un'identificazione errata, convincendolo che sia un esperto umano.

Edward Feigenbaum



Cos'è l'AI?

Altre varianti del test

Test del minimo segnale intelligente (1997):
Rappresenta un'astrazione massima del Test di Turing, concentrandosi esclusivamente sulla capacità di pensiero e limitando le risposte alle sole opzioni binarie (vero/falso o sì/no). Questo approccio mira a eliminare i problemi associati al linguaggio. Le domande sono quindi isolate, rendendo il test più simile a un test del QI che a una conversazione. Di solito viene utilizzato per raccogliere dati statistici utili per valutare le prestazioni dei programmi di intelligenza artificiale.

Chris McKinstry

Cos’è l’AI?

Altre varianti del test

Test di Turing inverso: Nel test di Turing inverso viene messo alla prova un computer (che assume il ruolo di intervistatore) per verificare se è in grado di determinare se sta interagendo con un essere umano o con un’altra macchina.

Cos’è l’AI?

Altre varianti del test

CAPTCHA: Si tratta di una forma di test di Turing inverso in cui, prima di poter eseguire un'azione su un sito Web, all'utente vengono presentati caratteri alfanumerici in un'immagine grafica distorta (o un altro rebus) e gli viene chiesto di digitarli. Questo ha lo scopo di impedire che i sistemi automatici vengano utilizzati per abusare del sito. La logica è che attualmente non esiste un software sufficientemente sofisticato da leggere e trascrivere l'immagine distorta (o comunque non è disponibile per l'utente medio), quindi se un sistema è in grado di farlo è probabile che sia un essere umano.

Cos’è l’AI?

Altre varianti del test

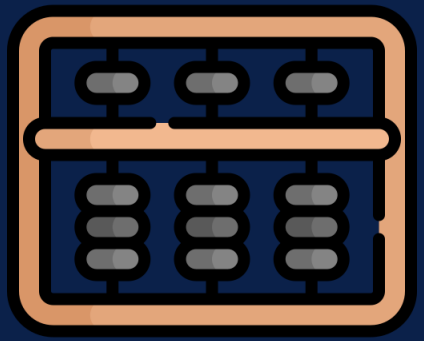
Test di Ebert (2011): È progettato per valutare se una voce sintetizzata al computer possiede abilità sufficienti, come intonazione, inflessione, tempismo, e altro ancora, per suscitare risate nelle persone.

Un po’ di storia

Storia dell’AI

Andrea Caruso

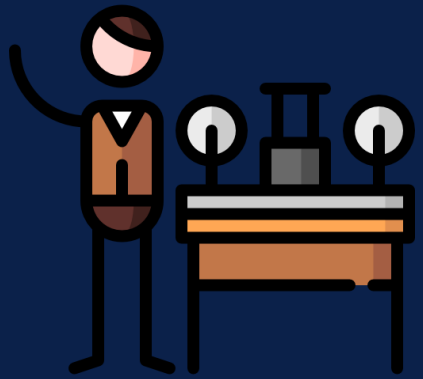
La tradizione di ricerca nell'intelligenza artificiale è stata influenzata da una serie di sviluppi storici. Iniziando con i primi tentativi di creare macchine calcolatrici nel 17° secolo, come quelle di Wilhelm Schickard e Blaise Pascal, fino alle idee di Gottfried Wilhelm von Leibniz per una macchina calcolatrice più avanzata. Nel 1830, Charles Babbage ha progettato la "macchina analitica", un precursore dei moderni calcolatori, mentre nel 1937 Claude Shannon ha mostrato il potenziale delle operazioni binarie per l'innovazione digitale.



Storia dell'AI

Andrea Caruso

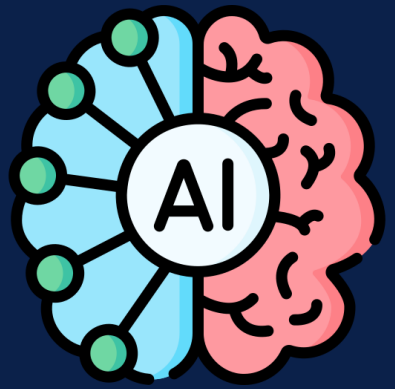
Un'altra pietra miliare è stata l'articolo di Alan Turing del 1936, che ha introdotto concetti fondamentali come la calcolabilità e la macchina di Turing. Nel 1943, McCulloch e Pitts hanno creato il primo lavoro rilevante sull'intelligenza artificiale, proponendo un modello di neuroni artificiali che può rappresentare funzioni computabili e connettivi logici. Nel 1950, Marvin Minsky e Dean Edmonds hanno sviluppato la prima rete neurale artificiale, chiamata SNARC. Questi sono solo alcuni degli sviluppi chiave che hanno plasmato la tradizione di ricerca nell'intelligenza artificiale.



Andrea Caruso

Storia dell’AI

Nel 1956, al Dartmouth College (Hanover, New Hampshire, U.S.A.) si tenne un convegno fondamentale nel campo della computazione dedicato allo sviluppo di sistemi intelligenti. John McCarthy, Marvin Minsky, Claude Shannon e Nathaniel Rochester parteciparono e proposero iniziative cruciali. McCarthy istituì un progetto ambizioso per creare una macchina in grado di simulare l'intelligenza umana, coinvolgendo ricercatori come Trenchard More, Arthur Samuel, Ray Solomonoff e Oliver Selfridge. Durante l'evento, Allen Newell e Herbert Simon presentarono il loro programma Logic Theorist. McCarthy coniò il termine "intelligenza artificiale", segnando la nascita di questa disciplina.



Storia dell'AI

Negli anni '50, Allen Newell e Herbert Simon crearono il General Problem Solver (GPS), un programma che imitava i processi di risoluzione dei problemi umani. Questo approccio differiva da quello del Logic Theorist (LP) di Newell e Simon. Nel frattempo, presso IBM, Nathaniel Rochester e colleghi svilupparono programmi di ragionamento simili. Nel 1959, Herbert Gelernter creò il Geometry Theorem Prover, mentre John McCarthy, al MIT, contribuì al campo definendo il linguaggio di programmazione Lisp e sviluppando l'idea di un "Advice Taker", un sistema intelligente completo progettato per trovare soluzioni a problemi non strettamente matematici. Marvin Minsky, sempre al MIT, coordinò la creazione di programmi per risolvere problemi limitati, noti come "micro mondi", tra cui il programma SAINT di James Slagle per il calcolo integrale.

Storia dell'AI

Andrea Caruso

Nel periodo successivo, i ricercatori nutrivano l'aspirazione di creare macchine con capacità di ragionamento simili a quelle umane. Herbert Simon, nel 1957, previde che entro dieci anni sarebbero state sviluppate macchine in grado di competere con i campioni di scacchi, anche se questa previsione si avverò solo dopo quarant'anni. Tuttavia, le sfide erano molte: la mancanza di conoscenza semantica nei domini trattati dalle macchine e la limitazione del ragionamento alla manipolazione sintattica causarono difficoltà. Nel 1966, il governo degli Stati Uniti interruppe i fondi per lo sviluppo delle macchine traduttrici a causa di queste difficoltà.

Inoltre, ci si rese conto che avere un algoritmo teoricamente efficace non garantiva la sua praticità nell'implementazione pratica. Infine, le limitazioni dei calcolatori nel rappresentare la logica furono evidenziate nel documento di Minsky e Papert del 1969, intitolato "Perceptrons", che dimostrò che anche un percettrone, una forma semplice di rete neurale, aveva limitazioni nel riconoscimento di pattern complessi.

Storia dell’AI

Le difficoltà incontrate nel campo dell'intelligenza artificiale portarono a definire gli approcci delle macchine come "approcci deboli" (slide successiva), richiedendo una conoscenza più approfondita nel campo di applicazione. Nel 1969, il programma DENDRAL, sviluppato da Ed Feigenbaum, Bruce Buchanam e Joshua Lederberg, rappresentò una svolta, poiché poteva ricostruire la struttura di una molecola a partire dai dati di uno spettrometro. DENDRAL fu il precursore dei sistemi che incorporarono i concetti teorizzati da McCarthy per l'Advice Taker. Successivamente, Feigenbaum e altri ricercatori di Stanford (California, U.S.A.) fondarono l'Heuristic Program Project (HPP) per estendere l'applicazione di questi sistemi, sviluppando il sistema MYCIN per la diagnosi delle infezioni sanguigne. Questi sistemi, noti come "sistemi esperti", possedevano una conoscenza specifica in determinati contesti, con l'uomo che trasferiva direttamente la propria conoscenza alla macchina attraverso regole logiche.

Storia dell'AI

Andrea Caruso

IA FORTE

L'intelligenza artificiale forte o generale (in sigla AGI dall'inglese Artificial General Intelligence) è la capacità di un agente intelligente di apprendere e capire un qualsiasi compito intellettuale che può imparare un essere umano.

IA DEBOLE

L'intelligenza artificiale debole o ristretta ha lo scopo di essere in grado di risolvere esattamente un singolo problema. Sono "deboli" quei programmi che non mostrano di avere una coscienza o una "mente" equiparabile a quella umana.

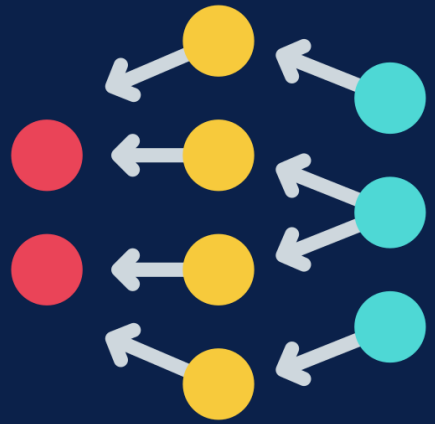
Andrea Caruso

Storia dell’AI

Negli anni '80, l'intelligenza artificiale trovò un'applicazione commerciale con il sistema R1 della Digital Equipment nel 1982, utilizzato per configurare ordini di nuovi computer. Nell’anno 1986, R1 risparmiò alla compagnia 40 milioni di dollari. Anche la DuPont risparmiò circa 10 milioni di dollari all'anno utilizzando sistemi simili. Negli Stati Uniti, quasi tutte le grandi aziende avevano sistemi esperti in funzione e stavano esplorando tecnologie più avanzate. In Giappone, nel 1981, fu annunciato il progetto Fifth Generation, con l'obiettivo di sviluppare sistemi intelligenti basati su Prolog, mentre negli Stati Uniti venne creata la Microelectronics and Computer Technology Corporation (MCC) per garantire la competitività nazionale nella ricerca. Nel Regno Unito, il rapporto Alvey recuperò i fondi tagliati dal rapporto Lighthill del 1973, che aveva interrotto il supporto alla ricerca in intelligenza artificiale. Tuttavia, molti di questi progetti non raggiunsero gli obiettivi previsti. Nel 1988, l'industria dell'intelligenza artificiale raggiunse un valore dell'ordine di miliardi di dollari, coinvolgendo centinaia di aziende nello sviluppo di sistemi esperti, robot e software e hardware specializzati.

Storia dell'AI

Negli anni '80, fu riproposto l'algoritmo di apprendimento per reti neurali chiamato back-propagation, originariamente concepito da Bryson e Ho nel 1969. Questo algoritmo trovò applicazione in vari problemi di apprendimento, sia informatici che psicologici. I modelli "connessionisti" basati su reti neurali emersero come alternative ai modelli simbolici proposti da Newell, Simon, McCarthy e altri, cercando di rispondere a domande che i modelli precedenti non erano riusciti a affrontare, sebbene incontrassero anch'essi alcune difficoltà. Di conseguenza, i due approcci, simbolico e connessionista, furono considerati complementari anziché esclusivi.



Andrea Caruso

Storia dell'AI

Oggi, i sistemi intelligenti sono pervasivi in ogni settore, comprese le attività quotidiane e i giochi, come previsto dagli esperti di intelligenza artificiale. Programmi come Deep Blue hanno affrontato con successo campioni di scacchi, mentre la NASA ha impiegato il programma Remote Agent per gestire le missioni spaziali nel 1998. Alcune auto sono ora dotate di sistemi di guida autonoma. Anche nelle routine quotidiane, come il controllo della temperatura con termostati intelligenti, i sistemi intelligenti interagiscono con l'ambiente e altri dispositivi. Nel settore economico, l'automazione e la tecnofinanza stanno rivoluzionando il tasso di occupazione e il panorama lavorativo in generale.

Andrea Caruso

Storia dell’AI

Python è il linguaggio dominante nell'intelligenza artificiale grazie alla sua velocità nel creare algoritmi complessi, rispetto ad altri linguaggi. La sua evoluzione ha portato alla versione 3, consolidandolo come standard nel campo. Il numero di librerie Python dedicate all'intelligenza artificiale, come PyTorch e Pandas, è cresciuto notevolmente nel tempo, contribuendo alla sua adozione diffusa.

Storia dell’AI

Andrea Caruso

Nel 2017, la Conferenza di Asilomar sulla IA Benefica (California) promossa dal Future of Life Institute, ha prodotto un vademecum con 23 principi per affrontare le questioni etiche, sociali, culturali e militari legate all'intelligenza artificiale. Questo documento ha ottenuto un ampio consenso, con la firma iniziale di oltre 800 esperti e il successivo sostegno di migliaia di altri.

Immagine Di Existential2021 - Opera propria, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=109870710

Andrea Caruso

Applicazioni dell’AI

Andrea Caruso

Applicazioni dell’AI

Modellizzazione predittiva

L'IA può essere utilizzata per costruire modelli predittivi che possono aiutare le aziende a prendere decisioni migliori. Ad esempio, l'apprendimento automatico può essere utilizzato per predire quali clienti sono più propensi ad acquistare un determinato prodotto, o quali pazienti sono più propensi a sviluppare una certa malattia.

Elaborazione del linguaggio naturale

L'IA è utilizzata per costruire sistemi che possono comprendere e interpretare il linguaggio umano. Questo è importante per applicazioni come il riconoscimento vocale, i chatbot e la traduzione linguistica.

Andrea Caruso

Applicazioni dell’AI

Computer Vision (visione artificiale)

L'IA è utilizzata per costruire sistemi che possono riconoscere e interpretare immagini e video. Questo è importante per applicazioni come le auto a guida autonoma, i sistemi di sorveglianza e l'imaging medico.

Industria sanitaria

Il settore sanitario è stato uno dei primi a utilizzare l'IA con il rilevamento delle immagini.

Applicazioni dell’AI

Andrea Caruso

Individuazione delle frodi

L'IA può essere utilizzata per rilevare comportamenti fraudolenti nelle transazioni finanziarie, nella pubblicità online e in altre aree.

Industria finanziaria

Le banche utilizzano principalmente l'IA per individuare modelli all'interno dei dati.

Organizzazione governativa

Il governo fa uso dell'IA per gestire la sicurezza pubblica e i servizi pubblici. Ad esempio la Cina fa massiccio uso del riconoscimento facciale.

Applicazioni dell’AI

Andrea Caruso

Automazione

I robot basati su IA eseguono fasi essenziali nei processi di produzione nelle fabbriche.

Trasporti

L'IA è utilizzata nell'industria dei trasporti per ottimizzare le rotte, ridurre il consumo di carburante e migliorare l'efficienza complessiva. Gioca anche un ruolo nei veicoli autonomi, dove gli algoritmi di apprendimento automatico vengono utilizzati per prendere decisioni sulla navigazione e sulla sicurezza.

Applicazioni dell’AI

Andrea Caruso

Sistemi di raccomandazione

L'IA può essere utilizzata per costruire sistemi di raccomandazione che suggeriscono prodotti, servizi o contenuti agli utenti in base al loro comportamento e preferenze passate.

Marketing

Prima dell'era dei big data, i ricercatori sviluppavano strumenti matematici avanzati come l'analisi bayesiana per stimare il valore di un cliente. Adesso i dipartimenti marketing si affidano all'IA.

Industria al dettaglio

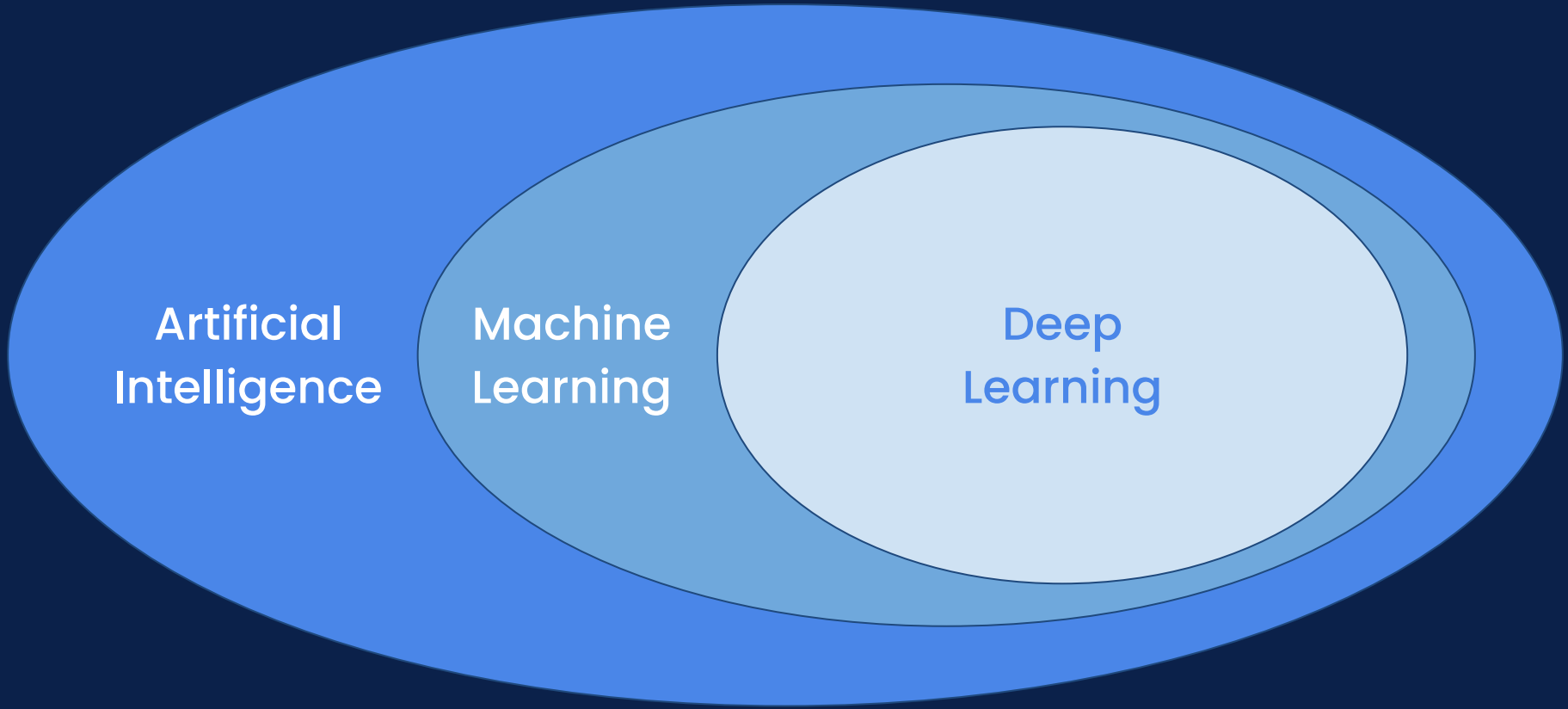
Grazie all’AI, si analizza il comportamento dei clienti, si prevede la domanda e si gestisce l'inventario di conseguenza. Aiuta anche i rivenditori a personalizzare l'esperienza di acquisto per ogni cliente raccomandando prodotti basati sui loro acquisti passati e sulle preferenze.

Andrea Caruso

Settori dell’AI

Settori dell'AI

Andrea Caruso



Andrea Caruso

Settori dell’AI

Intelligenza Artificiale (AI)

Si riferisce alla simulazione di processi di intelligenza umana da parte di computer e sistemi. Include una vasta gamma di tecniche e approcci per creare sistemi in grado di eseguire compiti che richiedono intelligenza umana, come l'apprendimento, il ragionamento, la comprensione del linguaggio e la visione artificiale.

Settori dell’AI

Machine Learning (Apprendimento automatico)

È una sottodisciplina dell'Intelligenza Artificiale che si concentra sullo sviluppo di algoritmi e modelli che consentono ai computer di apprendere da dati e di migliorare le proprie prestazioni nel tempo senza essere esplicitamente programmato per farlo. Il Machine Learning può utilizzare una varietà di tecniche, tra cui le reti neurali artificiali, per l'apprendimento dai dati.

Andrea Caruso

Settori dell’AI

Deep Learning (Apprendimento profondo)

È una sotto-categoria del Machine Learning che utilizza reti neurali artificiali profonde, ossia con molti livelli, per apprendere rappresentazioni gerarchiche dei dati. È particolarmente efficace nell'analisi di dati non strutturati, come immagini, testo e suoni, ed è responsabile dei recenti avanzamenti in settori come la visione artificiale (computer vision) e il riconoscimento del linguaggio naturale.

Andrea Caruso

Machine Learning

Machine Learning

Andrea Caruso

Il Machine Learning (ML) è una sotto-area dell'Intelligenza Artificiale (IA) che si concentra sullo sviluppo di sistemi che imparano o migliorano le prestazioni basandosi sui dati che elaborano.
Il ML conferisce ai computer la capacità di apprendere senza essere esplicitamente programmati. Ad oggi il ML viene utilizzato in molti più contesti di quanto ci si possa aspettare.

Machine Learning

Andrea Caruso

Caratteristiche generali del Machine Learning

- È una tecnologia basata sui dati, organizzati generalmente in dataset.
- Grandi quantità di dati vengono generati su base quotidiana. Individuando relazioni significative e pattern nei dati, si possono prendere decisioni opportune. In questo senso, è simile e spesso correlato al data mining.
- Le macchine possono imparare autonomamente dai dati e migliorare in autonomia.

Machine Learning

Andrea Caruso

Tipologie di problemi di machine learning

Esistono vari modi per classificare i problemi di machine learning.
In base alla natura del "segnale" o del "feedback" di apprendimento disponibile per un sistema di apprendimento, distinguiamo:

- Apprendimento supervisionato
- Apprendimento non supervisionato
- Apprendimento semi-supervisionato
- Apprendimento per rinforzo

Machine Learning

Andrea Caruso

Apprendimento supervisionato

All'inizio del processo, i dati vengono divisi in training set (dati di addestramento, solitamente il 70%-80%) e test set (dati di test, solitamente il 30%-20%).
AI sistema vengono presentati gli input e i relativi output desiderati (labels -> etichette), generalmente organizzati in un dataset. Il sistema trova pattern e connessioni tra l'input e l'output.
L'obiettivo è apprendere una regola generale che mappa gli input agli output. Il processo di addestramento continua fino a quando il modello raggiunge il livello desiderato di accuratezza non solo sui dati di addestramento, ma anche su nuovi dati mai visti prima (test set).

Alcuni esempi reali sono:

- Classificazione delle immagini
- Previsioni/Regressioni di mercato

Machine Learning

Andrea Caruso

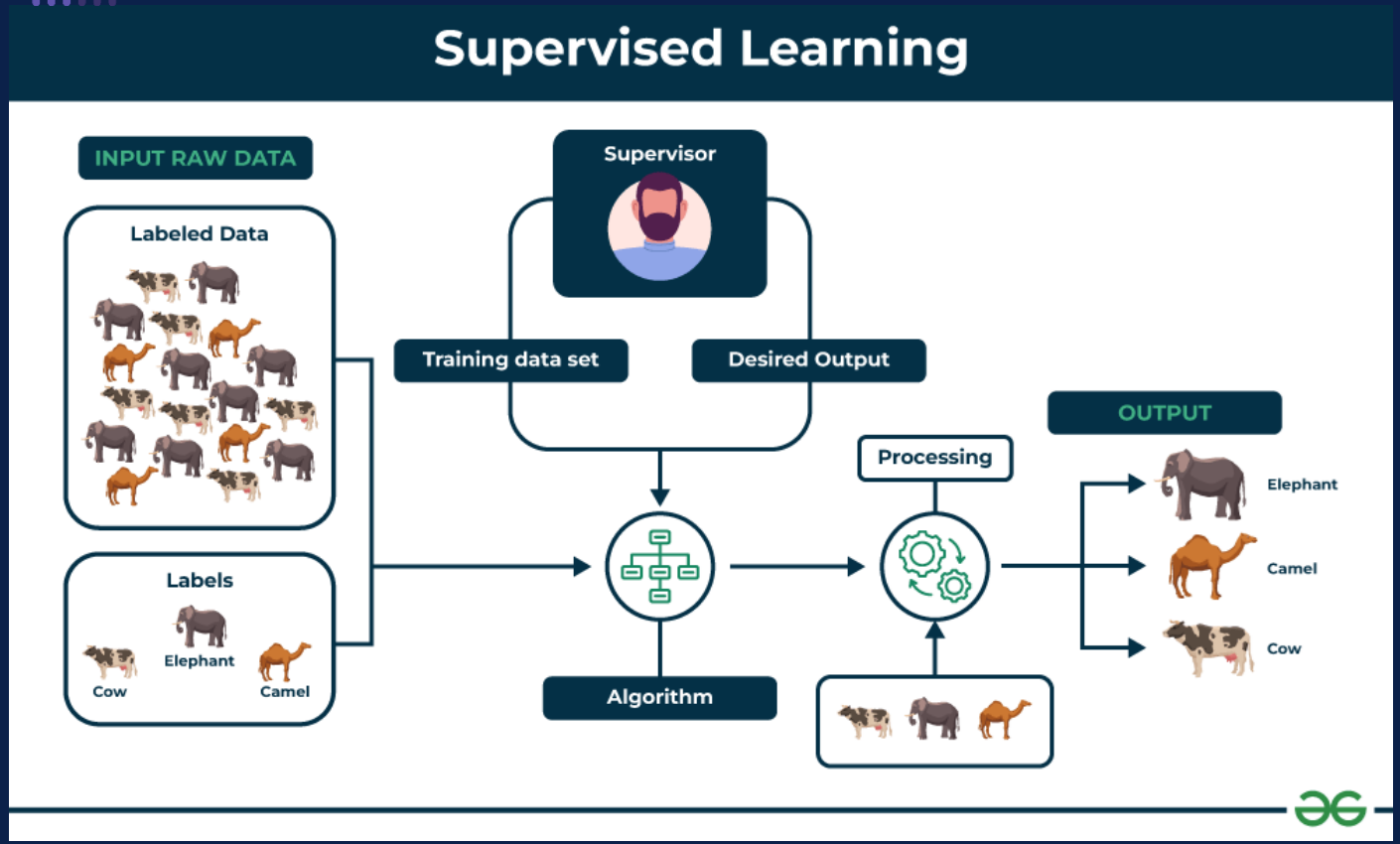


Immagine geeksforgeeks.org

Machine Learning

Andrea Caruso

Classificazione

Gli input vengono divisi in due o più classi, contrassegnate da etichette. Il sistema deve produrre un modello che assegna input su cui non è stato addestrato a una (classificazione binaria) o più (classificazione multi-etichetta) di queste classi.
Un esempio di classificazione binaria è il filtraggio dello spam, dove gli input sono messaggi di posta elettronica e le classi sono "spam" e "non spam", mentre un esempio di classificazione multi-etichetta è la rilevazione di eventuali masse tumorali, in cui si può distinguere tra “assenza”, “benigna”, “maligna”.

Regressione

Si addestra il sistema a prevedere valori numerici. L’addestramento avviene somministrando, ad esempio, dati di mercato storici, e si chiede al modello di predire il futuro prezzo di un’azione.

Machine Learning

Andrea Caruso

Un esempio di classificazione e regressione su due set di dati diversi è mostrato di seguito:

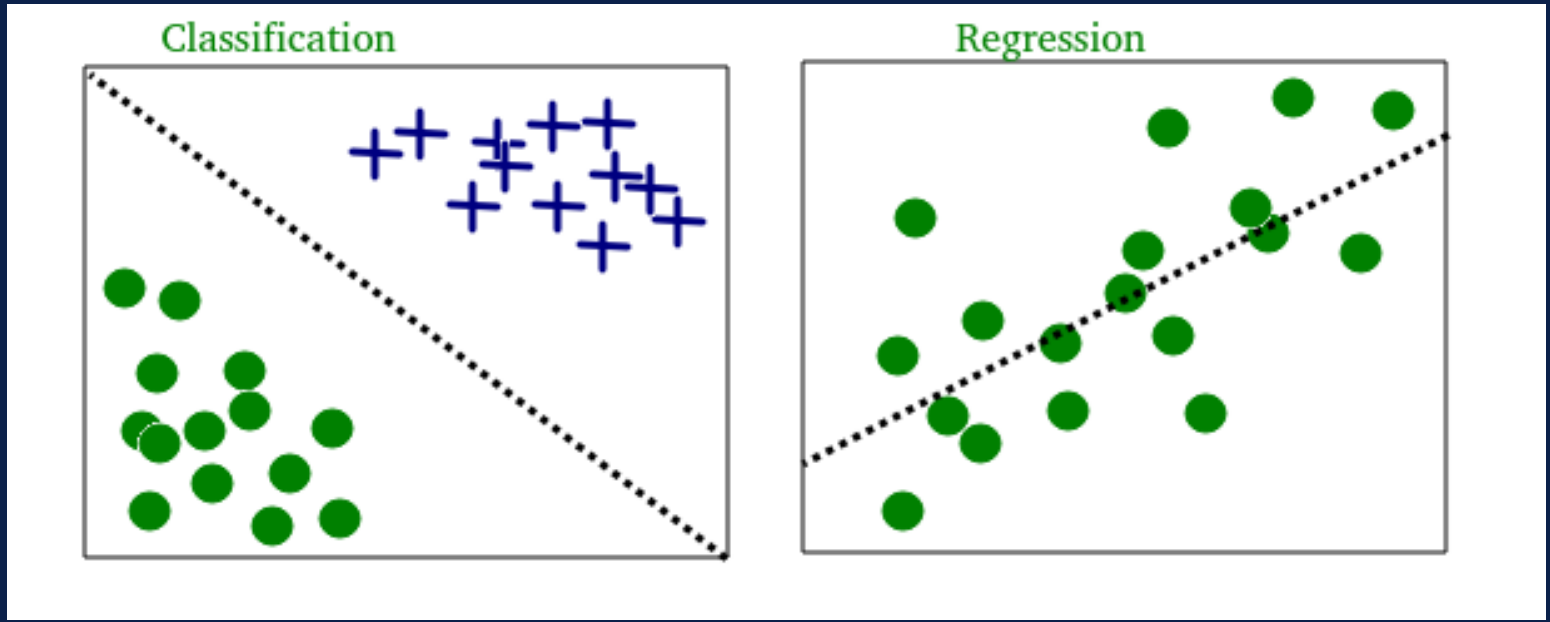


Immagine geeksforgeeks.org

Machine Learning

Andrea Caruso

Tra gli algoritmi di classificazione più utilizzati:

- Regressione Logistica
- Naive Bayes
- Alberi decisionali
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)

Tra quelli di regressione:

- Regressione Lineare
- Regressione Polinomiale
- Regressione Ridge
- Regressione ad albero Decisionale
- Regressione Random Forest
- Regressione Support Vector Machine (SVM)

Machine Learning

Andrea Caruso

Apprendimento supervisionato

Pro

I modelli di Supervised Learning possono avere un'alta precisione poiché vengono addestrati su dati etichettati.

Il processo decisionale nei modelli di supervised learning è spesso interpretabile.

Spesso può essere utilizzato mediante modelli pre-addestrati, il che risparmia tempo e risorse nello sviluppo di nuovi modelli.

Contro

Ha limitazioni nel riconoscere pattern e potrebbe avere difficoltà con pattern non visti o imprevisti che non sono presenti nei dati di addestramento.

Può essere dispendioso in termini di tempo e costi poiché si basa solo su dati etichettati.

Potrebbe portare a cattive generalizzazioni basate sui dati.

Machine Learning

Andrea Caruso

Apprendimento non supervisionato

Anche in questo caso, solitamente il dataset viene diviso in training set e test set.
Non vengono fornite etichette (output) all'algoritmo di apprendimento, lasciandolo libero di trovare da solo connessioni e pattern nel suo input. Viene utilizzato per raggruppare dati in gruppi diversi. L'apprendimento non supervisionato può portare a scoprire pattern nascosti nei dati.
Esempi di utilizzo:

- Clustering
- Analisi ad alta dimensionalità
- Modelli generativi: Dopo che un sistema immagazzina dati e quindi ne calcola la distribuzione di probabilità, può essere in grado di generare ulteriori dati.

Machine Learning

Andrea Caruso

Unsupervised Learning

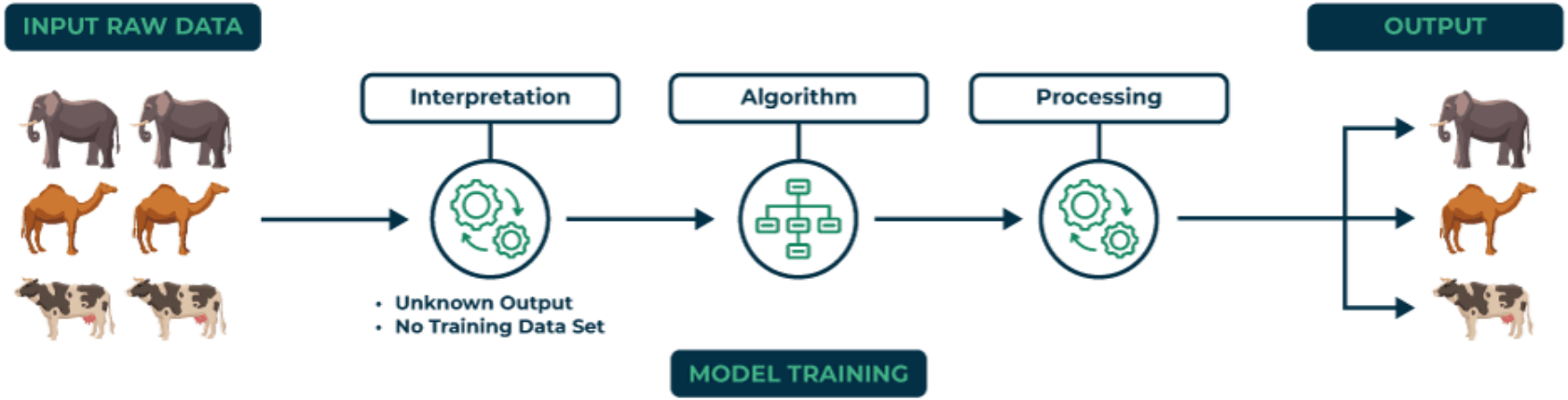


Immagine geeksforgeeks.org

Machine Learning

Andrea Caruso

Clustering

Si chiede al sistema di separare dati in gruppi distinti (cluster) per caratteristiche. A differenza della classificazione, non vi sono etichette note.

Analisi ad alta dimensionalità

Si utilizza il sistema per visualizzare e ispezionare dati a molte dimensioni.

Riduzione della dimensionalità

Semplifica gli input mappandoli in uno spazio a dimensione inferiore. Un esempio è il topic modeling, in cui un programma riceve un elenco di documenti in una determinata lingua (umana) e ha il compito di scoprire quali documenti trattano argomenti simili.

Modelli generativi

Dopo che un sistema ha immagazzinato dati e quindi ne ha calcolato la distribuzione di probabilità, può essere in grado di generare ulteriori dati.

Machine Learning

Andrea Caruso

Tra gli algoritmi di clustering più utilizzati:

- K-means
- Clustering gerarchico
- DBSCAN

Tra quelli di riduzione dimensionale:

- Principal Component Analysis (PCA)
- t-SNE
- Autoencoder

Machine Learning

Andrea Caruso

Apprendimento non supervisionato

Pro

Non richiede etichettatura, riducendo lo sforzo e i costi nella creazione del dataset.

Aiuta a scoprire pattern nascosti e varie relazioni tra i dati.

Contro

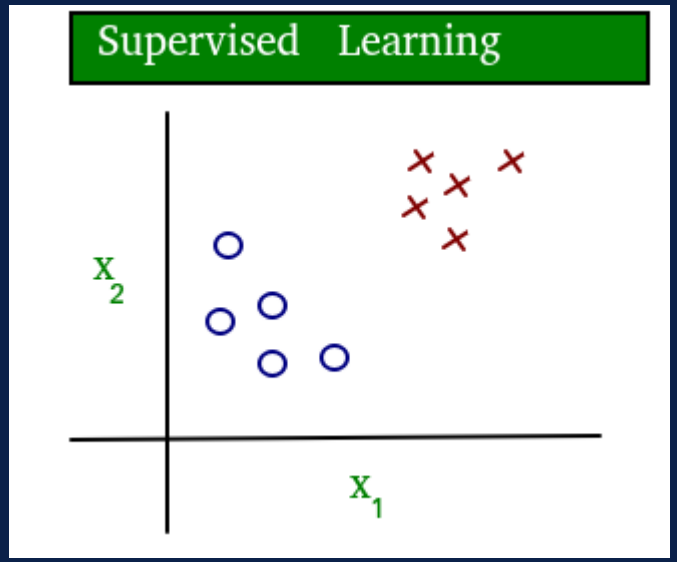
Senza l'uso di etichette, è spesso difficile prevedere la qualità dell'output.

L'interpretabilità dei cluster potrebbe non essere chiara o potrebbe addirittura non avere spiegazioni significative.

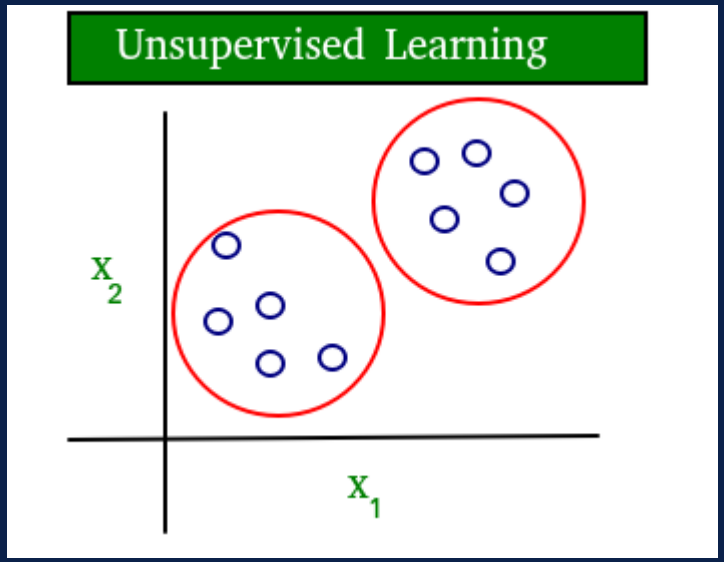
Machine Learning

Andrea Caruso

Apprendimento supervisionato



Apprendimento non supervisionato

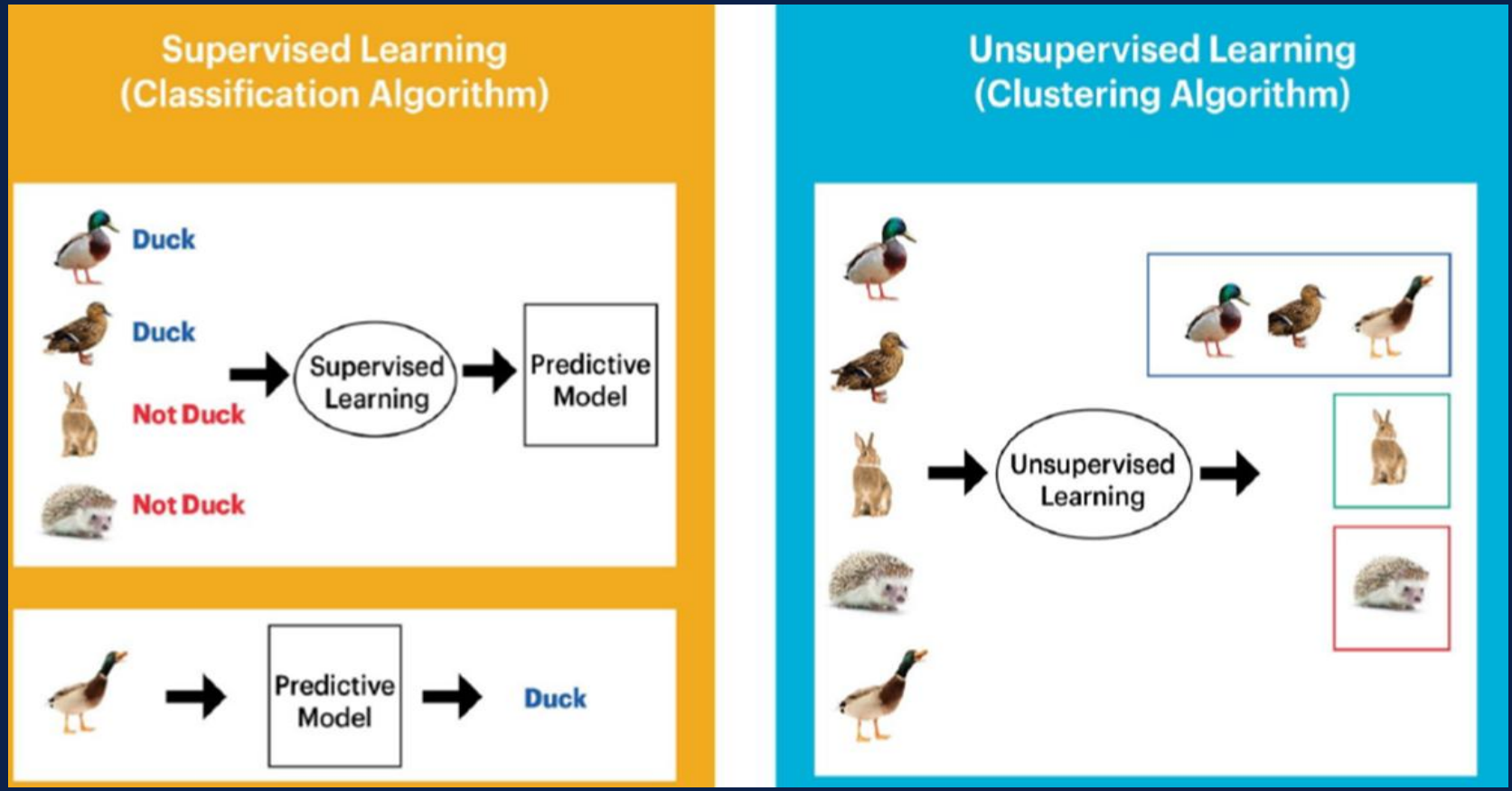


Si nota chiaramente che i dati nell'apprendimento supervisionato sono etichettati, mentre i dati nell'apprendimento non supervisionato sono non etichettati.

Immagini geeksforgeeks.org

Machine Learning

Andrea Caruso



Machine Learning

Apprendimento semi-supervisionato

I problemi in cui si ha una grande quantità di dati di input e solo alcuni di essi sono etichettati sono chiamati problemi di apprendimento semi-supervisionato. Questi problemi si collocano tra l'apprendimento supervisionato e quello non supervisionato. Ad esempio, un archivio fotografico in cui solo alcune immagini sono etichettate (ad esempio, cane, gatto, persona) e la maggior parte non lo è.

Machine Learning

Andrea Caruso

Apprendimento semi-supervisionato

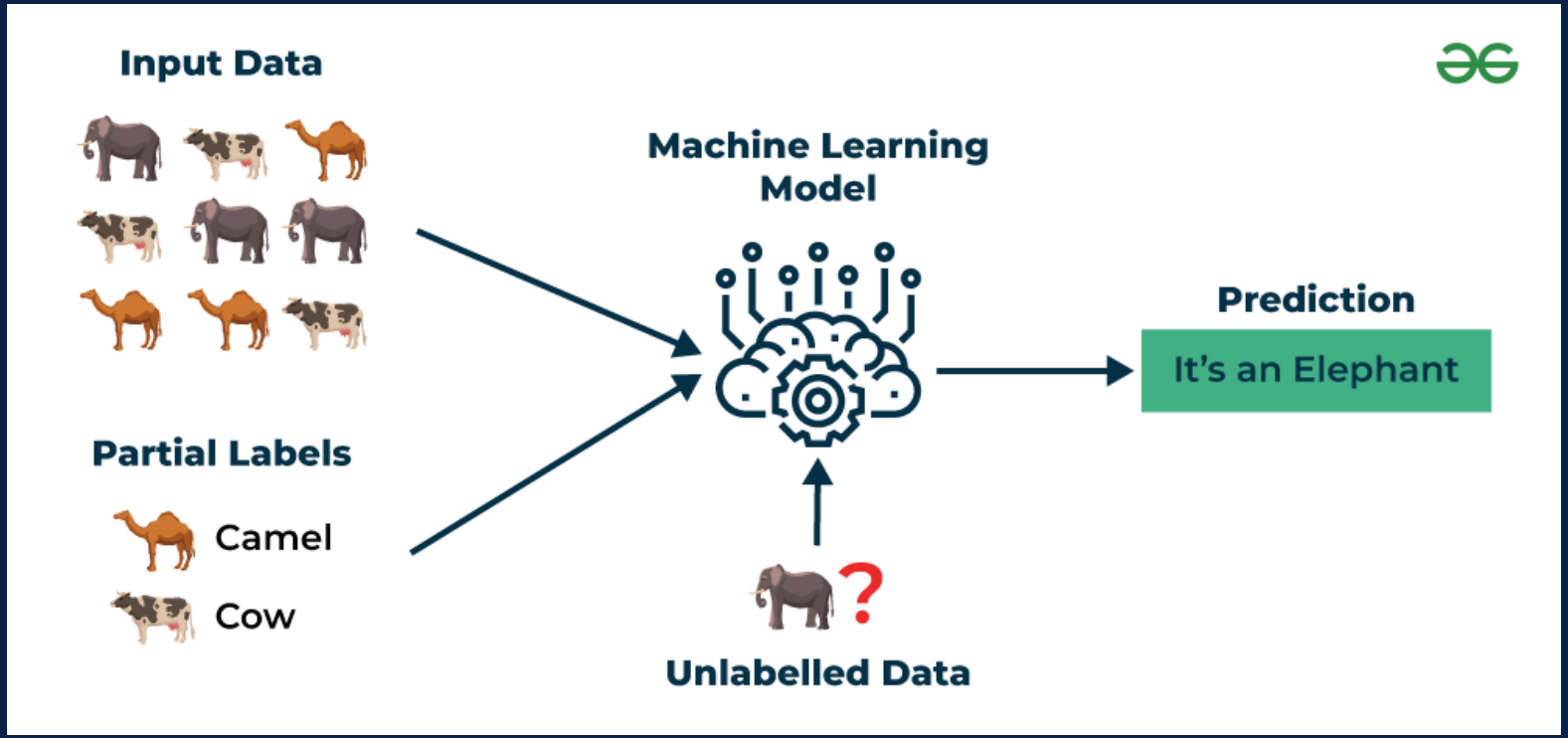


Immagine geeksforgeeks.org

Machine Learning

Apprendimento semi-supervisionato

Pro

Migliore generalizzazione rispetto al machine learning supervisionato, poiché utilizza sia dati etichettati che non etichettati.

Può essere applicato a una vasta gamma di dati.

Contro

I metodi semi-supervisionati sono in genere più complessi da implementare rispetto ad altri approcci.

Richiede comunque alcuni dati etichettati che potrebbero non essere sempre disponibili o facili da ottenere.

I dati non etichettati possono influenzare le prestazioni del modello di conseguenza.

Machine Learning

Andrea Caruso

Apprendimento per rinforzo

Il sistema interagisce con un ambiente dinamico in cui deve raggiungere un determinato obiettivo (come guidare un veicolo o sconfiggere un avversario in un determinato gioco). Il programma riceve feedback in termini di ricompense e punizioni mentre esplora le possibili scelte. L'obiettivo finale, in termini più tecnici, è quello di imparare una policy, che è una mappatura da stati ad azioni, che massimizza la ricompensa cumulativa attesa (expected cumulative reward) nel tempo.

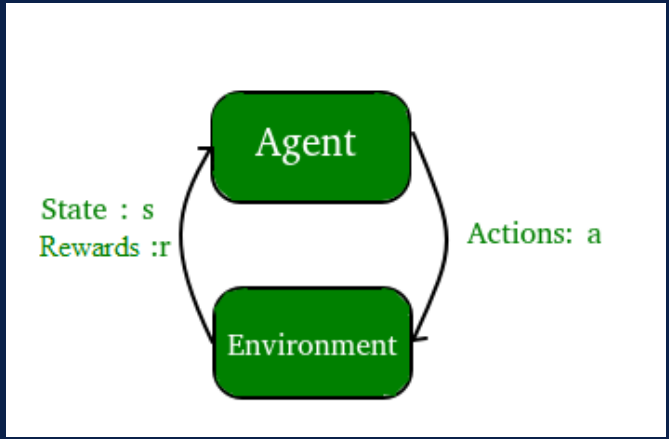
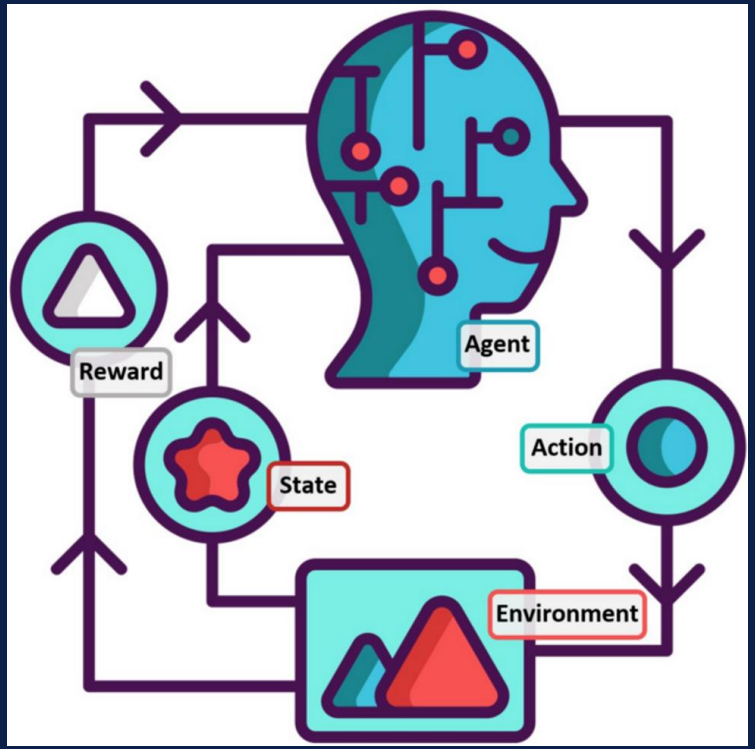


Immagine geeksforgeeks.org

Machine Learning

Andrea Caruso

Apprendimento per rinforzo



Machine Learning

Andrea Caruso

Distinguiamo due principali tipi di reinforcement learning:

- Reinforcement learning basato sul modello: L'agente costruisce un modello dell'ambiente, inclusa la probabilità di transizione tra gli stati e le ricompense associate a ciascuna coppia stato-azione. L'agente utilizza poi questo modello per pianificare le sue azioni al fine di massimizzare la ricompensa attesa. Alcuni popolari algoritmi di reinforcement learning basati sul modello includono Value Iteration e Policy Iteration.

- Reinforcement learning senza modello: In questo tipo di reinforcement learning, l'agente impara una policy direttamente dall'esperienza senza costruire esplicitamente un modello dell'ambiente. L'agente interagisce con l'ambiente e aggiorna la sua policy in base alle ricompense che riceve. Alcuni popolari algoritmi di reinforcement learning senza modello includono Q-Learning, SARSA e Deep Reinforcement Learning.

Machine Learning

Andrea Caruso

Alcuni algoritmi di reinforcement learning:

- Q-learning
- SARSA (State-Action-State-Reward-State-Action)
- Deep Q-learning

Andrea Caruso

Terminologia del ML

Andrea Caruso

Terminologia del ML

Model (Modello)

Un modello è una rappresentazione specifica appresa dai dati applicando un algoritmo di ML. Un modello viene anche detto ipotesi.

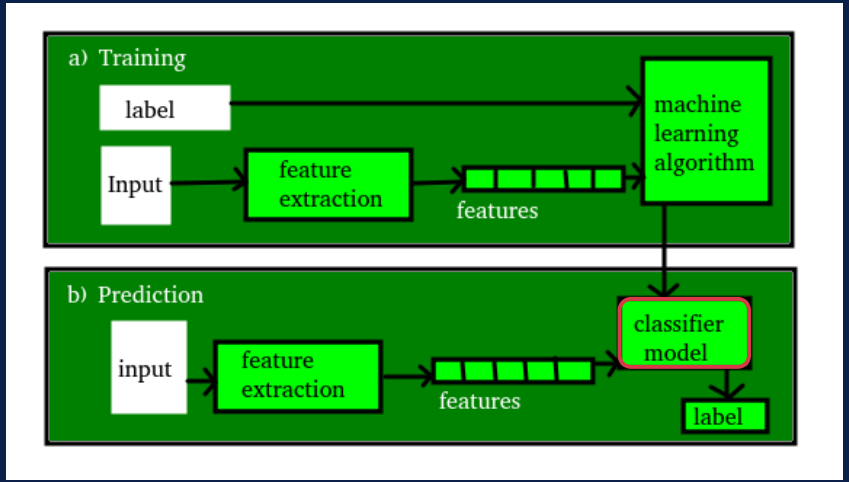


Immagine geeksforgeeks.org

Terminologia del ML

Andrea Caruso

Feature (Caratteristica)
Una caratteristica è una proprietà misurabile dei dati. Un insieme di caratteristiche numeriche può essere descritto da un vettore di dati. I feature vectors vengono forniti come input al sistema. Ad esempio, per prevedere la tipologia di un frutto, potrebbero esserci caratteristiche come colore, odore, sapore, ecc. Nota: La scelta di caratteristiche informative, discriminanti e indipendenti è un passo cruciale. Generalmente si utilizzano estrattori di features per ricavare caratteristiche rilevanti dai dati grezzi.

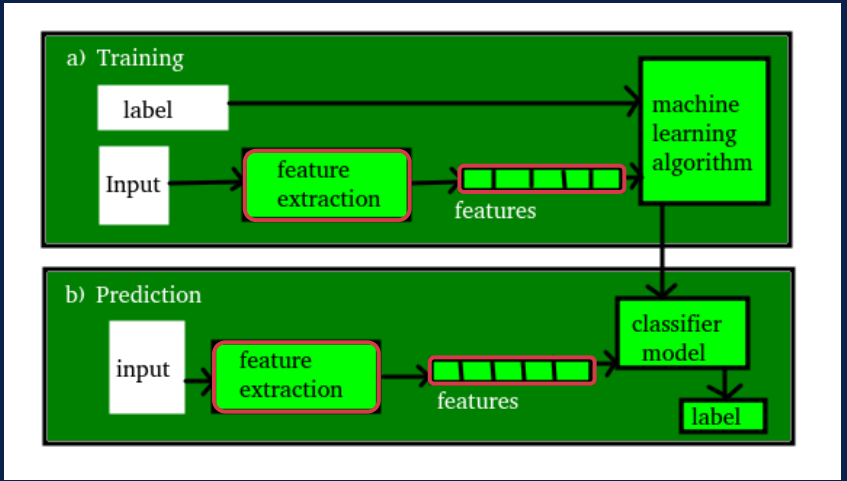


Immagine geeksforgeeks.org

Terminologia del ML

Andrea Caruso

Label (etichetta) o target variable (variabile obiettivo)
Un'etichetta o variabile target è il valore che il modello dovrebbe essere in grado di prevedere una volta addestrato. Nel ML supervisionato (e parzialmente in quello semi-supervisionato), vengono fornite insieme ai dati grezzi in fase di addestramento. Riprendendo l'esempio del frutto, l'etichetta è il nome del frutto: "mela", "arancia", "banana", ecc.

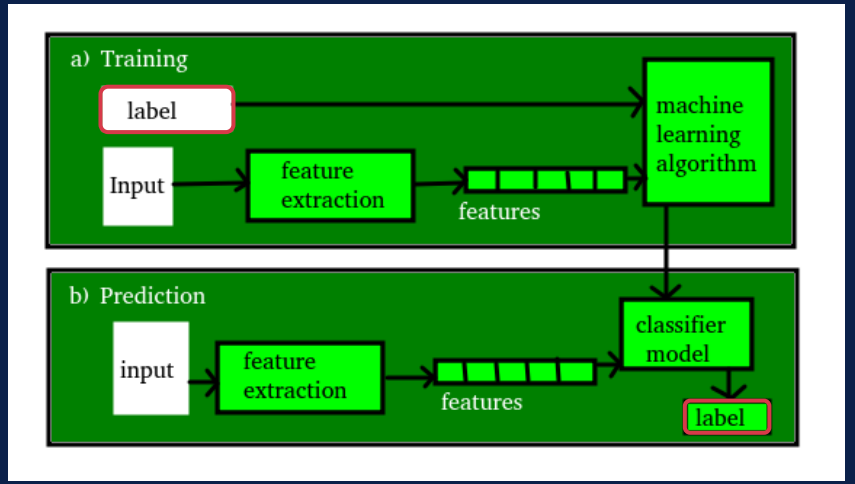


Immagine geeksforgeeks.org

Terminologia del ML

Andrea Caruso

Training (Addestramento)
L'idea è quella di fornire un insieme di input (features -> caratteristiche) e i relativi output attesi (labels -> etichette), quindi dopo l'addestramento, avremo un modello (ipotesi) che mappi i nuovi dati in una delle categorie addestrate.

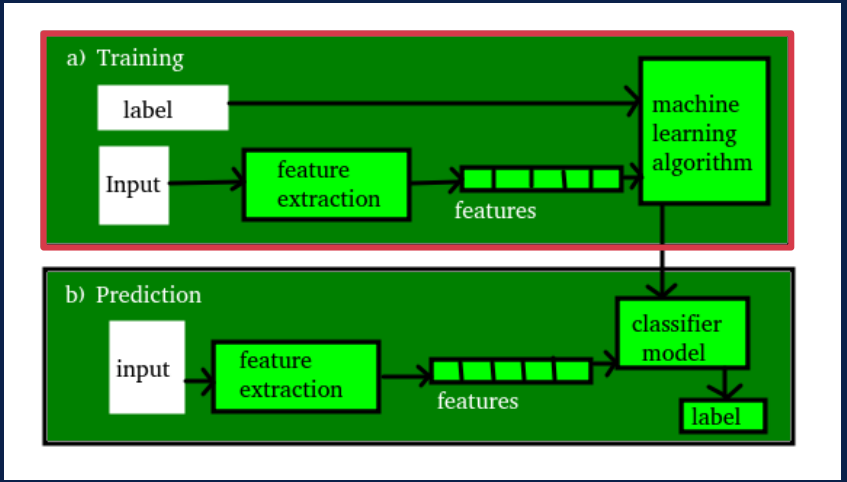


Immagine geeksforgeeks.org

Terminologia del ML

Prediction (Predizione)
Una volta che il modello è pronto, può essere alimentato con un insieme di input mai visti prima, dai quali fornirà degli output (etichette). Se il sistema riesce a prevedere correttamente gli output su dati su cui non è stato addestrato, entro un certo target di accuratezza, allora il sistema funziona correttamente.

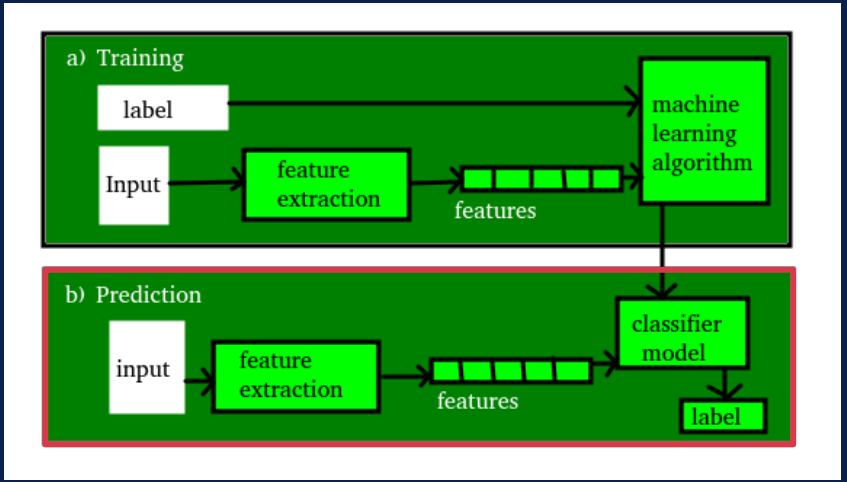


Immagine geeksforgeeks.org

Andrea Caruso

Step del ML

Step del ML

Andrea Caruso

1) Definire il problema: Identificare il problema che si vuole risolvere e determinare se il ML è la strada migliore per risolverlo.

2) Gestire i dati

a. Raccogliere i dati: Raccogliere i dati che verranno utilizzati per addestrare il modello. La qualità del modello dipenderà dalla qualità dei dati.

b. Esplorare i dati: Utilizzare la visualizzazione dei dati e i metodi statistici per comprendere la struttura e le relazioni all'interno dei dati.

c. Pre-elaborare i dati: Preparare i dati per la modellazione normalizzandoli, trasformandoli e ripulendoli se necessario.

d. Dividere i dati: Dividere i dati in training set e test set.

Step del ML

Andrea Caruso

3) Scegliere un Modello: Selezionare il modello di ML appropriato per il problema e i dati raccolti.

4) Addestrare il Modello: Utilizzare il training set per addestrare il modello, regolando i suoi parametri per adattarsi ai dati il più accuratamente possibile.

5) Valutare il Modello: Utilizzare il test set per valutare le prestazioni del modello e determinare la sua accuratezza.

6) Raffinare il Modello: Sulla base dei risultati della valutazione, raffinare il modello regolando i suoi parametri e ripetendo il processo di addestramento fino a quando non si raggiunge il livello desiderato di accuratezza.

Step del ML

Andrea Caruso

7) Distribuire/Integrare il Modello: Integrare il modello nella propria applicazione o sistema, rendendolo eventualmente disponibile per l'uso da parte di altri.

8) Monitorare il Modello: Monitorare continuamente le prestazioni del modello per garantire che continui a fornire risultati accurati nel tempo.

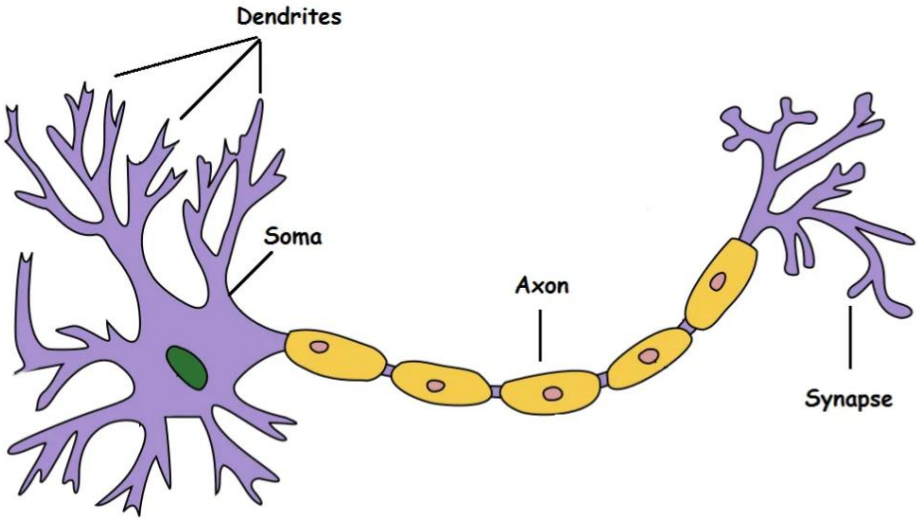
Andrea Caruso

Reti neurali

Reti neurali

Andrea Caruso

È ben noto che l'unità fondamentale delle reti neurali è chiamata percettrone o neurone artificiale. Il primo passo verso il percettrone che usiamo oggi è stato compiuto nel 1943 da McCulloch e Pitts, imitando la funzionalità di un neurone biologico.



Dendrite: Riceve segnali da altri neuroni

Soma: Elabora le informazioni

Assone: Trasmette l'output del neurone

Sinapsi: Punto di connessione ad altri neuroni

Reti neurali

Andrea Caruso

I nostri organi di senso interagiscono con il mondo esterno e inviano le informazioni visive e uditive ai neuroni. Supponiamo che si stia guardando un telefilm comico. Le informazioni che il cervello riceve vengono elaborate dal gruppo di neuroni "ridere o non ridere" che decideranno se attivare o meno la risata. Ogni neurone viene attivato solo quando il suo criterio di attivazione viene soddisfatto.

Ovviamente, nella realtà, non si tratta di un paio di neuroni che prendono decisioni. Nel nostro cervello esiste una rete interconnessa in modo massicciamente parallelo di 10^{11} neuroni (100 miliardi), e le loro connessioni non sono affatto semplici da comprendere.

Andrea Caruso

Reti neurali

Gli organi di senso passano le informazioni al primo strato di neuroni per elaborarle e l'output di questo strato viene trasmesso agli strati successivi in modo gerarchico. Alcuni neuroni si attiveranno e altri no. Questo processo a catena continua fino a che non si ottiene una risposta finale. Nell'esempio di prima, un'eventuale risata.

Questa rete massicciamente parallela garantisce anche una suddivisione del lavoro. Ogni neurone si attiva solo quando il suo livello di attivazione viene soddisfatto.

Reti neurali

Andrea Caruso

Si ritiene che i neuroni siano disposti in modo gerarchico (anche se non tutti gli scienziati concordano) e che ogni strato abbia il suo ruolo e la sua responsabilità. Per svolgere un compito complesso, come rilevare un volto, il cervello potrebbe fare affidamento sull'intera rete o gran parte di essa e non su un singolo strato.

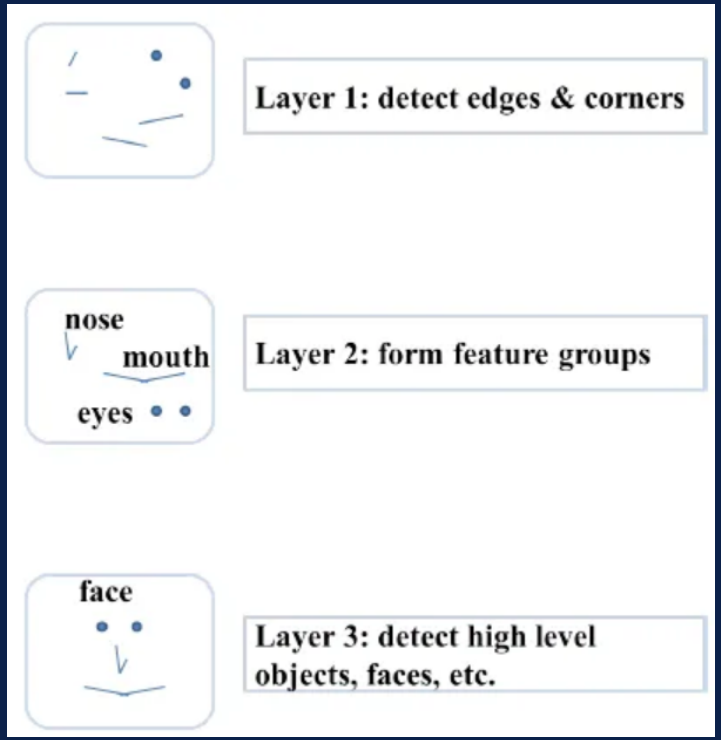
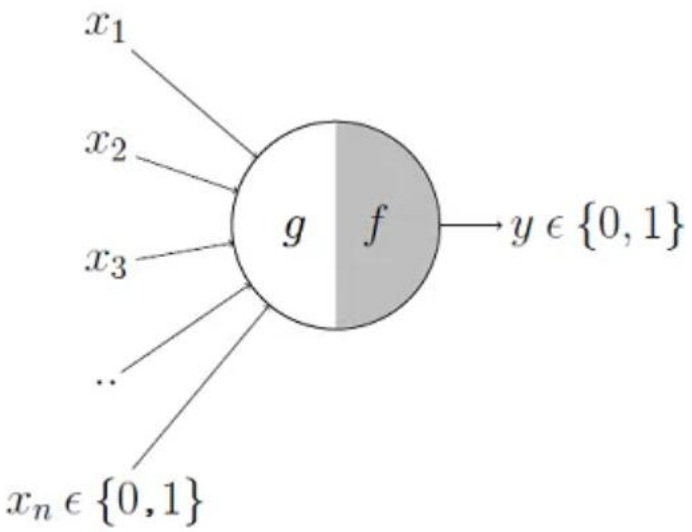


Immagine Mitesh M. Khapra's

Reti neurali

Andrea Caruso

Come detto, il primo modello computazionale di un neurone fu proposto da Warren McCulloch (neuroscienziato) e Walter Pitts (logico) nel 1943.



Semplificando, potremmo dividere il neurone in 2 parti.

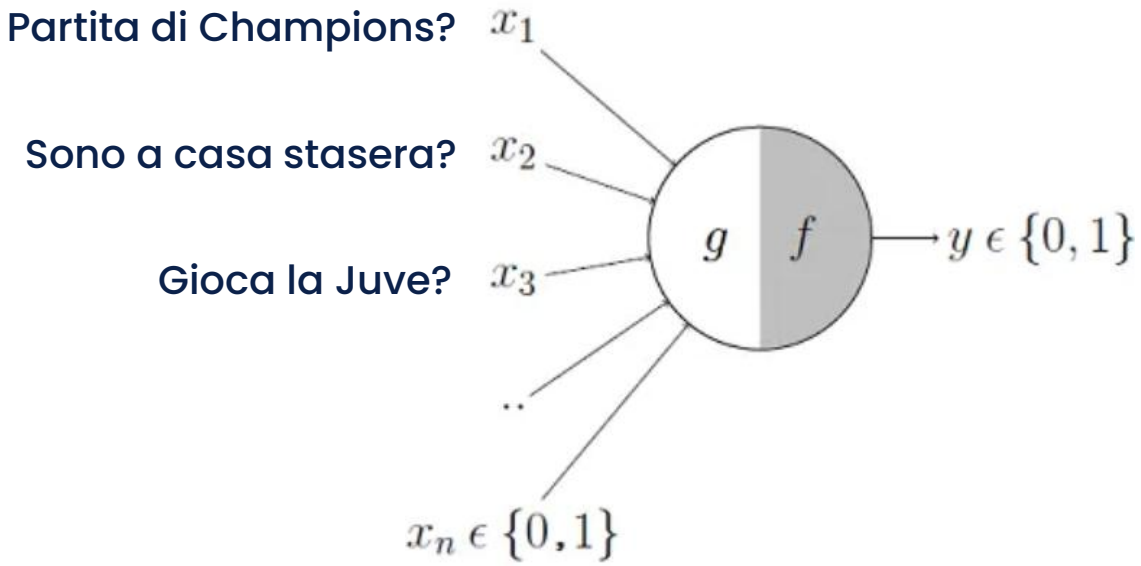
La parte di sinistra (g) riceve gli input e li aggrega.

In base al valore aggregato la parte di destra (f) si attiva o meno.

Reti neurali

Andrea Caruso

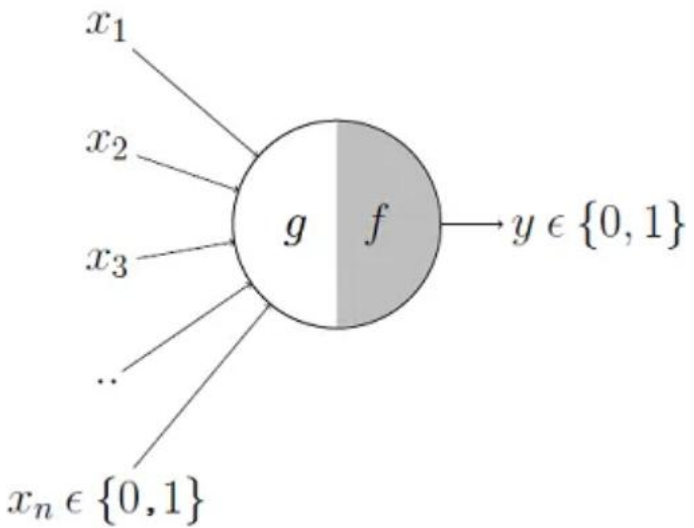
Supponiamo che voglia prevedere la mia decisione in merito a guardare o meno una partita di calcio. Gli input sono tutti booleani, cioè {0,1}, e anche la mia variabile di output è booleana {0: La guarderò, 1: Non la guarderò}.



Reti neurali

Andrea Caruso

Questi input possono essere sia eccitatori, sia inibitori. Si noti che g(x) sta semplicemente sommando gli input. E θ (theta) è il parametro di soglia, cioè il valore di g(x) che determina l'attivazione (1) o la non attivazione (0) del neurone.



$$g(x_1, x_2, x_3, ..., x_n) = g(\mathbf{x}) = \sum_{i=1}^{n} x_i$$

$$y = f(g(\mathbf{x})) = 1 \quad if \quad g(\mathbf{x}) \geq \theta$$
$$= 0 \quad if \quad g(\mathbf{x}) < \theta$$

Reti neurali

Andrea Caruso

Supponiamo di voler predire un valore di output y, dato un valore di input X. Ad esempio, si potrebbe voler prevedere il punteggio in un test universitario basato sul numero di ore di sonno e di studio del giorno (e della notte) precedente.

Per utilizzare un approccio di machine learning, abbiamo prima bisogno di dati.

X		y
Sonno	Studio	Voto
3	5	75
5	1	82
10	2	93

Supponiamo che per gli ultimi tre test un soggetto abbia registrato il numero di ore di studio, il numero di ore di sonno e il suo punteggio nel test. Utilizziamo Python per memorizzare i dati in array numpy bidimensionali.

Reti neurali

Andrea Caruso

Ora che abbiamo dei dati, li utilizzeremo per addestrare un modello a predire il voto al prossimo test, basandoci sul numero di ore di sonno e di studio. Questo è un problema di regressione supervisionata. È supervisionato perché i nostri dati hanno input e output. È un problema di regressione perché stiamo prevedendo il punteggio al test, che è un output continuo. Se stessimo prevedendo il voto finale (Ottimo - A / Distinto - B / ecc), sarebbe un problema di classificazione, e non un problema di regressione. Per risolvere il problema, utilizziamo una rete neurale artificiale.

	X		y
	Sonno	Studio	Voto
	3	5	75
	5	1	82
	10	2	93
	8	3	?

Reti neurali

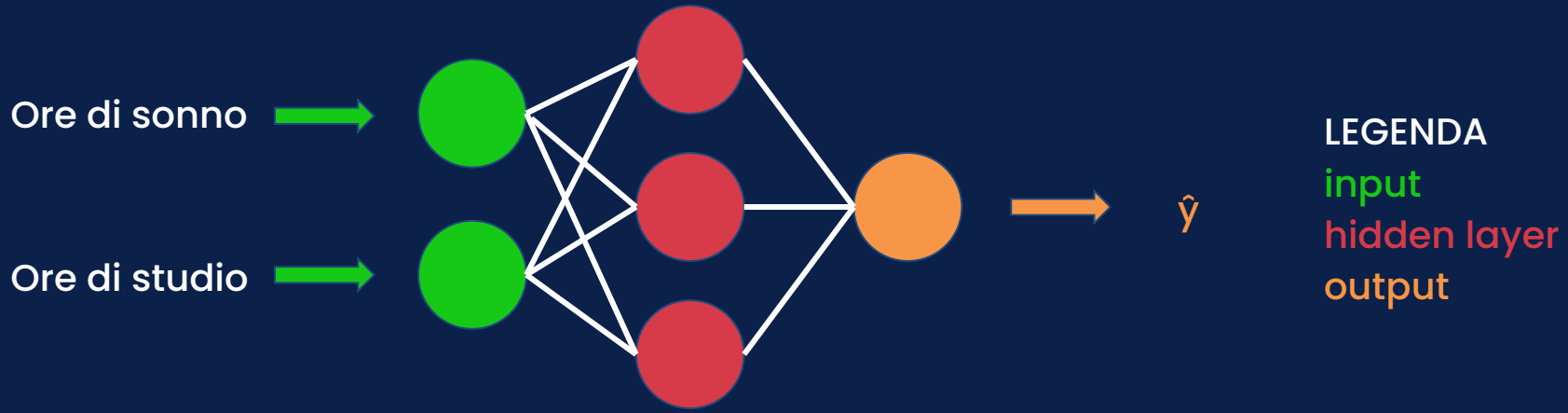
Andrea Caruso

Prima di inserire i dati nel modello, dobbiamo tener conto delle differenze nelle unità di misura. Entrambi gli input sono in ore, ma l'output è un punteggio scalato tra 0 e 100. Le reti neurali sono intelligenti, ma non abbastanza da azzeccare le unità di misura dei dati. La soluzione è scalare (normalizzare) i dati, in modo che il modello debba gestire solo unità standard, senza tener conto dell'effettiva unità di misura. Approfittiamo del fatto che tutti i dati sono positivi e semplicemente dividiamo ogni variabile per il valore massimo, scalando i valori tra 0 e 1.

Reti neurali

Andrea Caruso

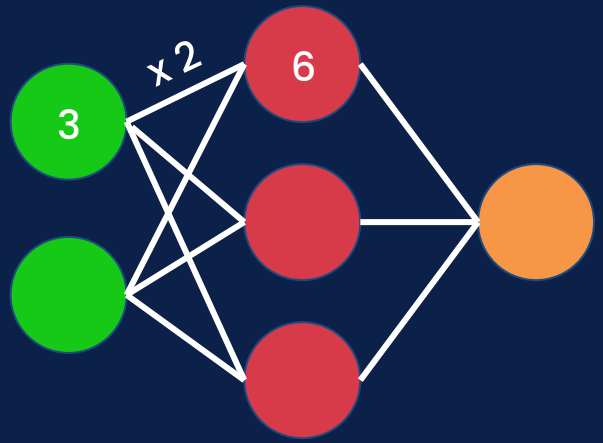
Costruiamo la nostra rete neurale. Questa deve avere 2 input e 1 output. Chiameremo l'output $\hat{y}$ (y hat), perché è una stima di Y, ma non coincide con Y. Qualsiasi strato tra quelli di input e di output è detto strato nascosto (hidden layer). Ad oggi, generalmente si costruiscono reti con molti strati nascosti dette deep networks, da cui deriva il termine deep learning. Qui utilizzeremo 1 strato nascosto con 3 nodi nascosti. Per ampliare la rete basta impilare insieme molti strati.



Reti neurali

Andrea Caruso

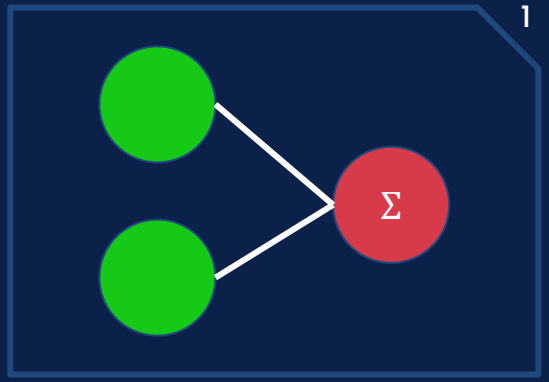
Nelle rappresentazioni visive delle reti neurali, i cerchi rappresentano i neuroni e le linee rappresentano le sinapsi (archi). Le sinapsi hanno un compito molto semplice, prendono un valore dal loro input, lo moltiplicano per un peso specifico e restituiscono il risultato.



Reti neurali

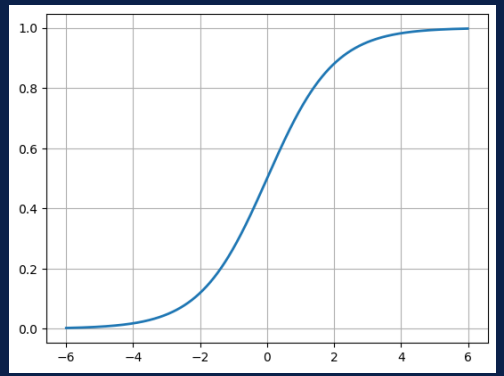
Andrea Caruso

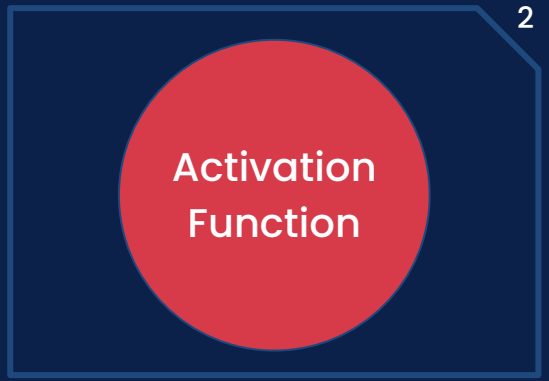
I neuroni sono un po' più complicati. Il loro compito è sommare insieme gli output di tutte le loro sinapsi e applicare una funzione di attivazione.
Alcune funzioni di attivazione consentono alle reti neurali di modellare pattern non lineari complessi, che modelli più semplici potrebbero non cogliere.



$$f(z) = \frac{1}{1 + e^{-z}}$$

e = Numero di Nepero = 2.71828

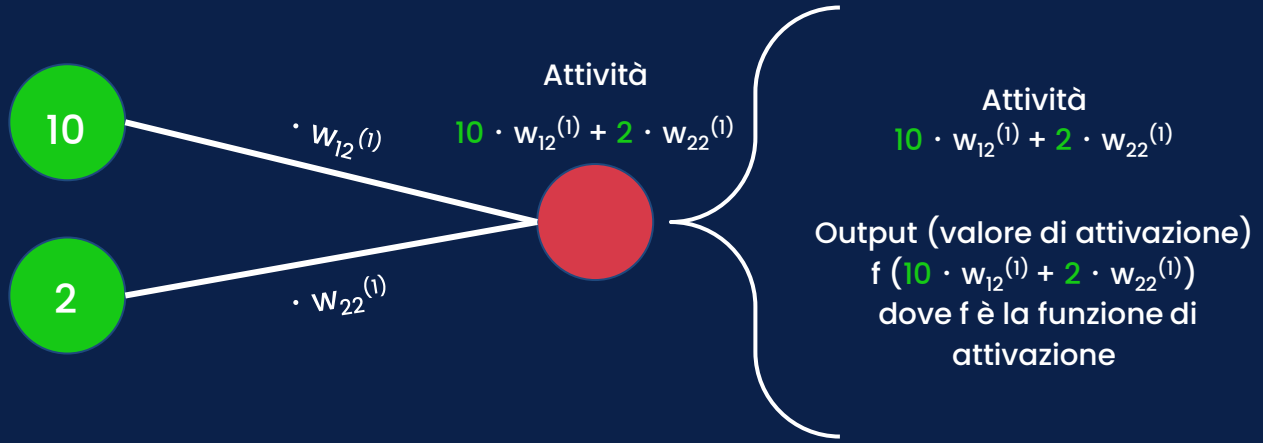




Reti neurali

Andrea Caruso

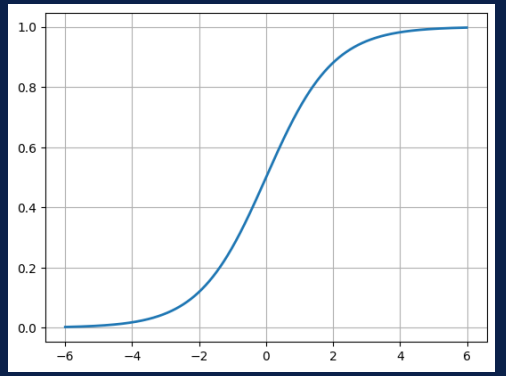
L'attività di un nodo in una rete neurale è la somma ponderata delle uscite dei nodi precedenti, ognuna moltiplicata per il peso del collegamento tra il nodo precedente e quello attuale. Questa somma ponderata è poi passata attraverso una funzione di attivazione per produrre l'output del nodo. L'attività di un nodo rappresenta quindi il contributo combinato dei nodi precedenti, mentre l'output del nodo è determinato dall'applicazione della funzione di attivazione all'attività. L'output è detto anche valore di attivazione.



Reti neurali

Andrea Caruso

Come funzione di attivazione, utilizzeremo la funzione di attivazione sigmoide.
La funzione sigmoide non ha una soglia di attivazione come alcune altre funzioni di attivazione, come ad esempio la funzione di attivazione step. La sigmoide produce un output compreso tra 0 e 1. La sigmoide trasforma l'input pesato attraverso una curva a forma di "S", dove a valori più alti di input corrispondono a valori di output più vicini a 1 e i valori più bassi di input corrispondono a valori di output più vicini a 0. L'output della sigmoide varia continuamente in base all'input, fornendo una rappresentazione graduale dell'attivazione del neurone.

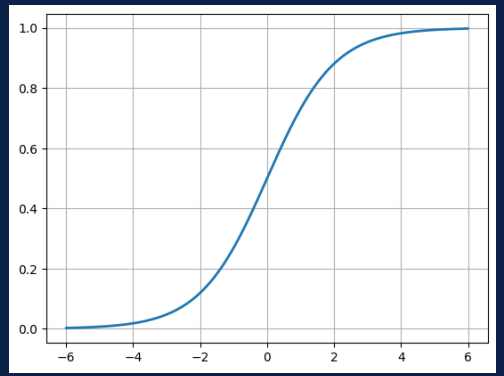


$$f(z) = \frac{1}{1 + e^{-z}}$$

e = Numero di Nepero = 2.71828

Reti neurali

Andrea Caruso



$$f(z) = \frac{1}{1 + e^{-z}}$$

e = Numero di Nepero = 2.71828

$$\text{sigmoide}(-5) = \frac{1}{1 + e^{5}}$$

Per z =-5 vale circa
0.006692850924284

$$\text{sigmoide}(1) = \frac{1}{1 + e^{-1}}$$

Per z =1 vale circa
0.7310585786300049

$$\text{sigmoide}(10) = \frac{1}{1 + e^{-10}}$$

Per z =10 vale circa
0.9999546021312976

Reti neurali

Andrea Caruso

Ma come si addestra una rete neurale?
Si utilizza un metodo detto retropropagazione dell'errore (backpropagation).
La retropropagazione dell'errore è un algoritmo per l'addestramento delle reti neurali artificiali, usato in combinazione con un metodo di ottimizzazione (ad esempio la discesa del gradiente).

La funzione di costo è una misura dell'errore tra le previsioni della rete e i valori attesi. È una funzione matematica che calcola quanto lontane sono le previsioni della rete dai risultati attesi. L'obiettivo dell'addestramento è minimizzare questa funzione di costo, in modo che le previsioni della rete siano il più vicine possibile ai risultati attesi.

La discesa del gradiente è una tecnica utilizzata nell'addestramento delle reti neurali per aggiornare i pesi della rete. Funziona calcolando il gradiente (derivata multidimensionale) della funzione di costo rispetto ai pesi della rete e quindi aggiornando i pesi in modo da muoversi nella direzione in cui la funzione di costo diminuisce. La discesa del gradiente aiuta quindi a trovare i valori dei pesi che minimizzano l'errore complessivo.

Reti neurali

Andrea Caruso

La retropropagazione richiede un'uscita desiderata per ogni valore in ingresso per poter calcolare il gradiente della funzione di costo.
Si tratta quindi di un metodo di apprendimento supervisionato, sebbene venga usato anche in reti non supervisionate. La retropropagazione richiede che la funzione d'attivazione usata dai neuroni artificiali (o "nodi") sia differenziabile.

Le fasi della retropropagazione dell'errore sono le seguenti:

- Forward Pass
- Calcolo dell'errore
- Backward Pass (retropropagazione)
- Aggiornamento dei pesi
- Ripetizione

Reti neurali

Andrea Caruso

- Forward Pass: Durante il passaggio in avanti, i dati di input vengono dati in pasto alla rete neurale che genera delle previsioni. Durante questo processo, vengono calcolati i valori di attivazione per ciascuno strato nascosto e di output.

- Calcolo dell'errore: Una volta ottenute le previsioni, si calcola l'errore confrontando le previsioni della rete con i valori di output desiderati (etichette reali). La funzione di costo viene utilizzata per misurare quanto l'output della rete si discosta dalle etichette reali.

- Backward Pass (Retropropagazione): Durante il passaggio all'indietro, si calcola il gradiente (derivata multidimensionale) della funzione di costo rispetto ai pesi della rete (cioè quanto varia la funzione di costo in relazione a ciascun peso) utilizzando la regola della catena. Questo viene fatto partendo dall'ultimo strato della rete e procedendo all'indietro verso il primo strato. Il gradiente viene calcolato iterativamente partendo dall'ultimo strato della rete e procedendo all'indietro verso il primo strato.

Reti neurali

- Aggiornamento dei pesi: Una volta calcolato il gradiente, si utilizza l'algoritmo di ottimizzazione (solitamente la discesa del gradiente) per aggiornare i pesi della rete in modo tale da ridurre gradualmente l'errore complessivo. Questo viene fatto spostandosi nella direzione opposta al gradiente, poiché vogliamo minimizzare la funzione di costo.

- Ripetizione: Il processo di forward pass, calcolo dell'errore, backward pass e aggiornamento dei pesi viene ripetuto per un numero fissato di epoche o fino a quando non si raggiunge una certa soglia di errore. Lo scopo finale è minimizzare l’errore complessivo, quindi minimizzare il valore della funzione di costo.