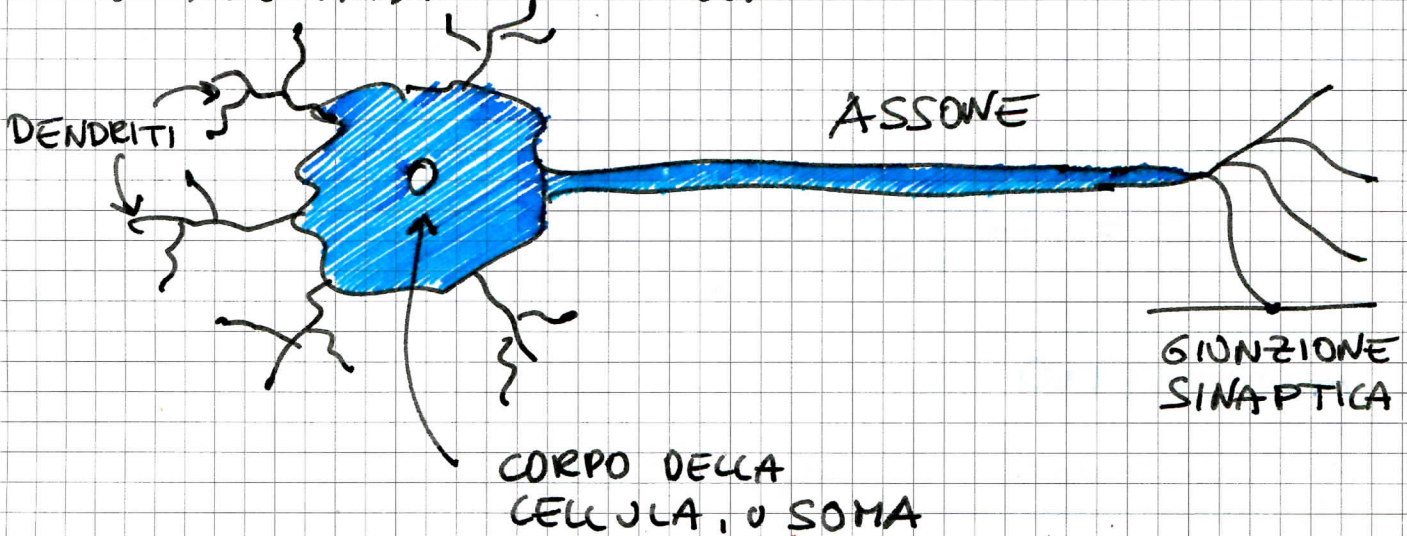


# RETI NEURALI

I modelli di reti neurali o neuronali sono degli algoritmi con funzioni cognitive come l'apprendimento e l'ottimizzazione di processi e sono basati su concetti derivati dallo studio sulla natura del cervello.

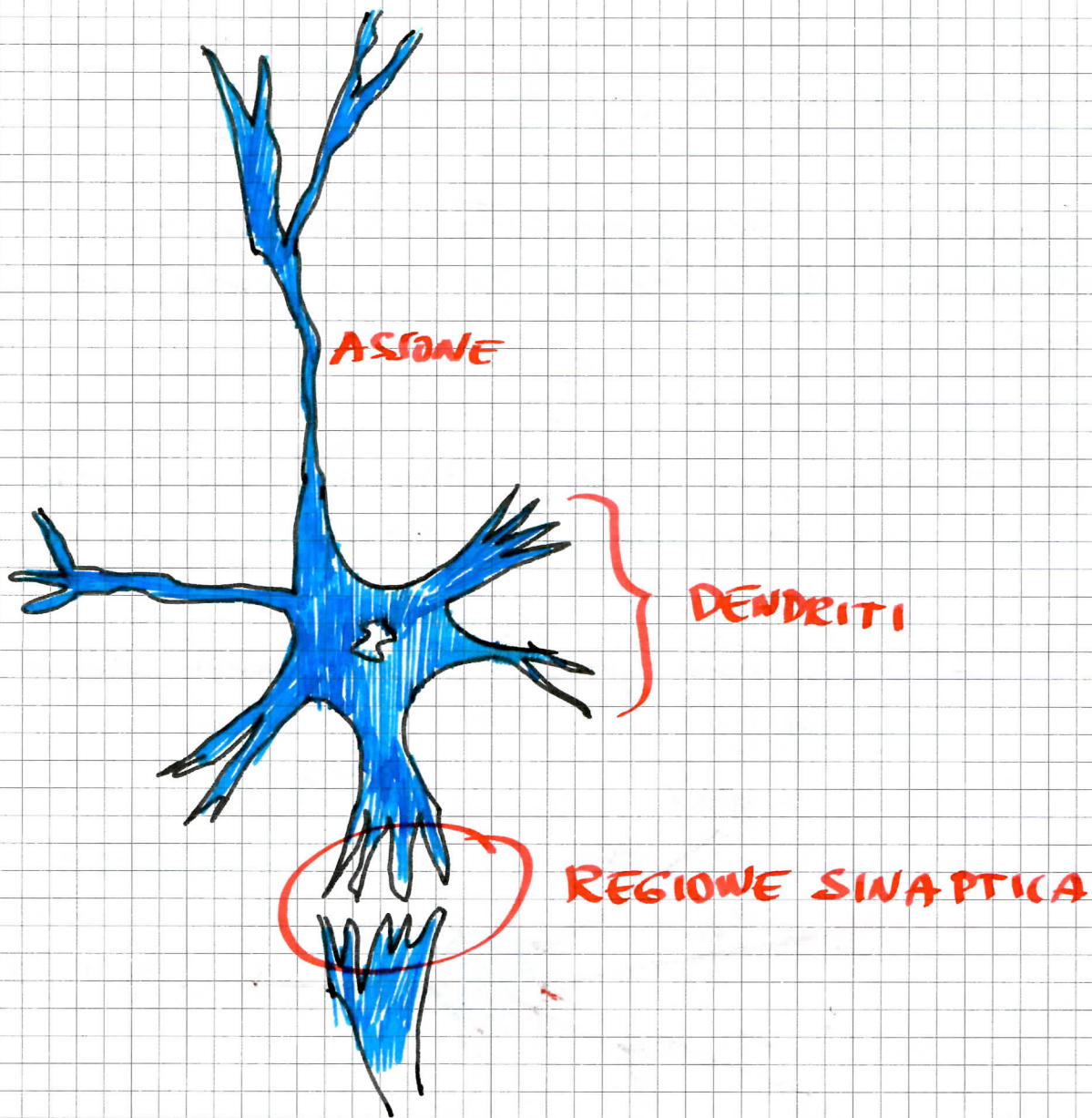
Il neurone è l'unità cellulare fondamentale di un sistema nervoso.



Ogni neurone ha una lunga fibra in uscita, chiamata ASSONE e delle terminazioni chiamate synapse. Dal corpo della cellula, detto anche SOMA, partono delle ramificazioni chiamate DENDRITI.

Ogni neurone riceve e combina i segnali che provengono da numerosi neuroni e ad essi collega tramite i dendriti. input da  
Un neurone riceve variate migliaia di connessioni provenienti da altri neuroni.





Dal punto di vista matematico, una rete neurale è definita come un grafo con le seguenti proprietà

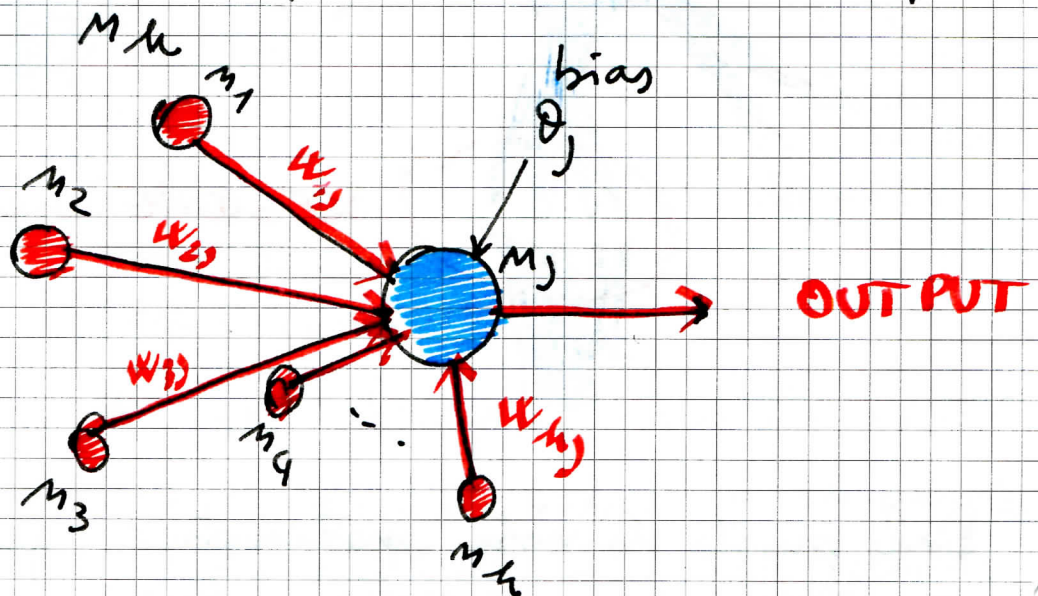
- 1) ad ogni nodo  $j$  è associata una variabile di stato  $n_j$
- 2) per ogni collegamento tra due neuroni  $i$  e  $j$  è definito un peso reale  $w_{ij}$
- 3) ad ogni nodo è associato un "bias" reale  $\theta_j$
- 4) è definita una funzione di trasferimento



$$f_j [m_h, w_{jh}, \theta_j, (h \neq j)]$$

La funzione di trasferimento determina l'evoluzione dello stato di un nodo ed è funzione:

- del bias  $\theta_j$ ,
- dei pesi  $w_{jh}$  associati ai link che collegano il neurone  $j$  con i neuroni  $h$
- dello stato presente dei nodi collegati,



Secondo la terminologia standard:

nodo  $\Rightarrow$  neurone

link  $\Rightarrow$  sinapi

bias  $\Rightarrow$  soglia di attivazione

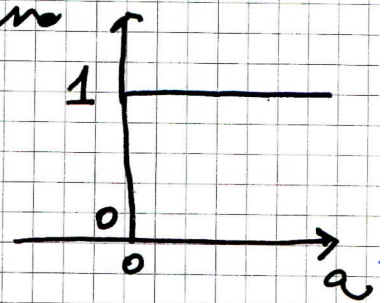
La funzione di trasferimento assume solitamente la forma

$$f_j \left( \sum_h w_{jh} m_h - \theta_j \right)$$



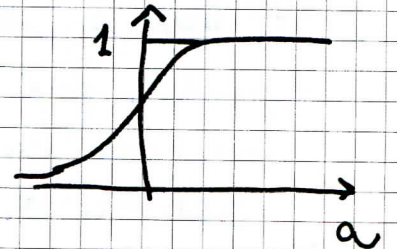
dove  $f$  è una funzione a gradino

$$f(a) = \begin{cases} 1 & a \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$



oppure una sigmoide

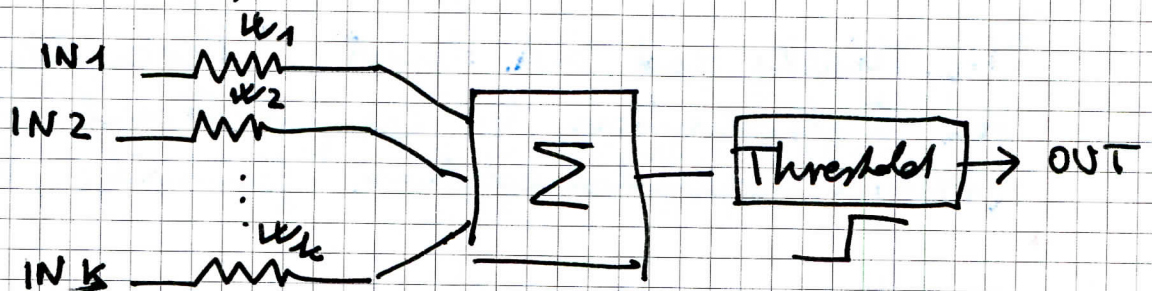
$$f(a) = \frac{1}{1 + e^{-a}}$$



Il primo modello di rete neurale è quello di McCulloch e Pitts del 1943 (MCP)

[W.S. McCulloch and W. Pitts, "A logical Calculus of the ideas Immanent in Nervous Activity", Bull. Math. Biophys. 5 (1943) 115.]

È un modello elettronico molto semplice che può essere implementato direttamente in hardware: un dispositivo che somma gli input a lui connessi (e 'pesati' fraente resistenze) e confronta il segnale ottenuto con una soglia, discriminando (o sopra soglia)

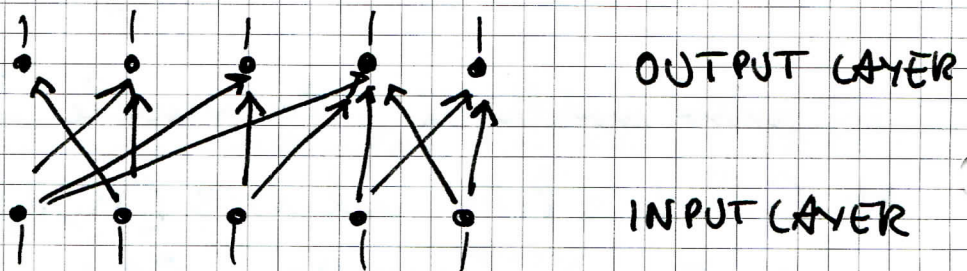


Alla fine degli anni '50, Rosenblatt e Wightmann introducono il Perceptron come evoluzione del MCP. Il sistema, nella sua forma più semplice, è costituito da due



strati di neuroni, uno di input e uno di output. Una prima applicazione della macchina Perceptron MARK I è nella pattern recognition.

F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain". *Psychological Review* 65 (1958), 386-408.



Il gruppo di Rosenblatt introduce anche un algoritmo iterativo per determinare i pesi delle connessioni.

Successivamente, Minsky e Papert dimostrano che un semplice perceptron non era in grado di risolvere problemi non linearmente separabili come realizzare, per esempio, la funzione XOR.

M. L. Minsky and S. A. Papert, "Perceptrons. An Introduction to Computational Geometry". (1988). Expanded edition. Cambridge, MA: MIT Press

Il risultato ~~che~~ ha bloccato congelando ogni sviluppo sulle reti neurali per quasi vent'anni.

Nel 1982, Hopfield pubblicò due articoli importanti che vennero considerati come la



viscosità delle vie nelle reti neurali.

Nel 1974 Little individua le sorprendenti similitudini tra il MCP ed un sistema di particelle di spin  $1/2$  su un reticolo, chiamato Modello di Ising:

- tutti i nodi hanno spin up ( $S_j = +1$ ) oppure spin down ( $S_j = -1$ )

L'analisi con il modello di Ising ha permesso di capire e studiare le proprietà termodinamiche del sistema

L'equazione deterministica

$$m_j(t+1) = f_j \left( \sum_k w_{jk} m_k - \theta_j \right)$$

viene sostituita da una espressione stocastica e lo stato di un nodo viene assegnato in base probabilistica.

Viene definita una funzione energia e la probabilità contiene un parametro  $T$  che è l'equivalente della temperatura del sistema fisico analogo.

W.A. Little and G.L. Shaw, "Analytic Study of the Memory Capacity of a Neural Network" Math. Biosci. 39, (1978), 281.

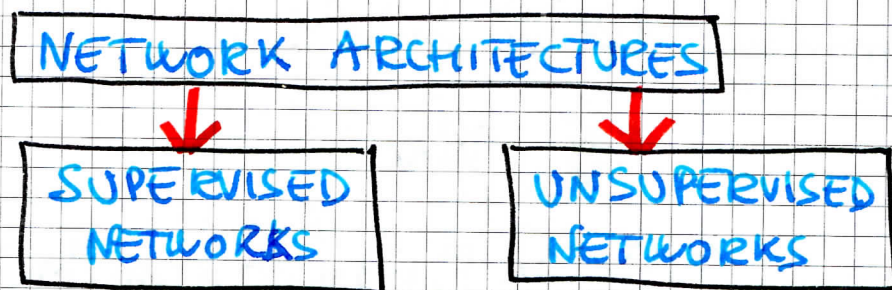
J.J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", Proc. Nat. Acad. Sci USA 79 (1982), 2554.

Le reti di Hopfield sono chiamate reti associative.



Al giorno d'oggi le reti neurali hanno grande applicazioni nella scienza che vanno al di là del semplice problema delle pattern recognition e si usano e studiano nelle fisica, ingegneria, finanza, scienze economiche e sociali, ecc.

Genericamente le architetture di reti neurali sono classificate in due categorie principali, supervised networks e unsupervised networks



La caratteristica saliente dei Supervised Networks è la presenza di un insegnante (teacher) o supervisore durante la fase di apprendimento.

Durante la fase di apprendimento, sono necessari coppie di input-output. Il vettore

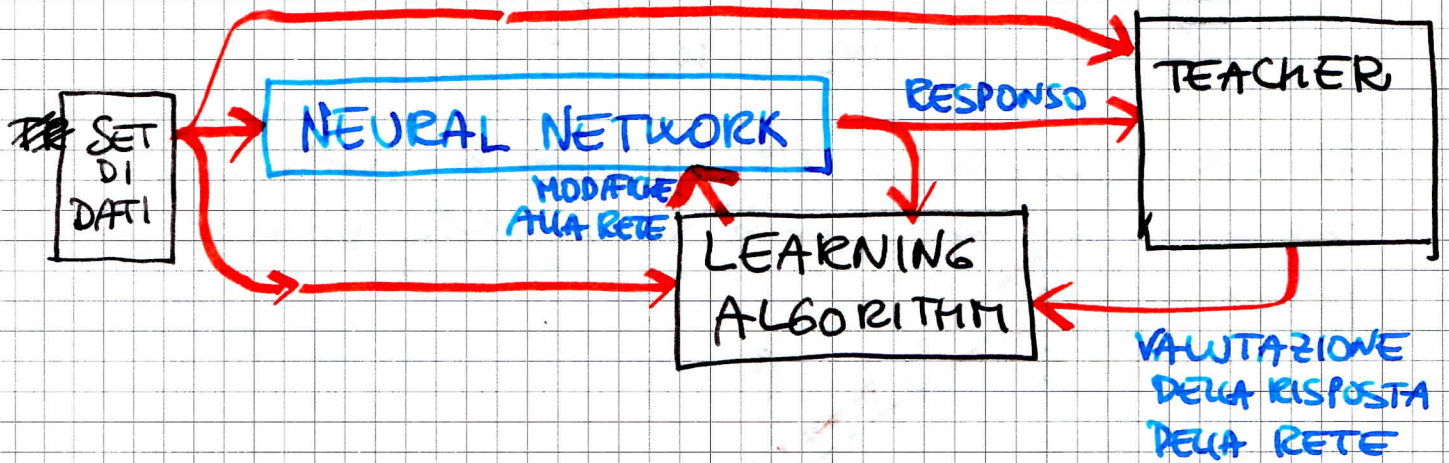
di output aspettati funziona come supervisore per valutare l'output fornito dalla rete nel processare i dati di input.

Durante la fase iterativa di apprendimento i pesi tra i neuroni della rete sono modificati in modo riprodurre, il più possibile, l'output desiderato. È possibile definire un segnale di errore come differenza tra output previsto e output prodotto dalla rete.



La procedura di apprendimento continua fino a quando gli errori sono vicini a zero o al di sotto di un valore di soglia predefinito.

Schema di SUPERVISED LEARNING:

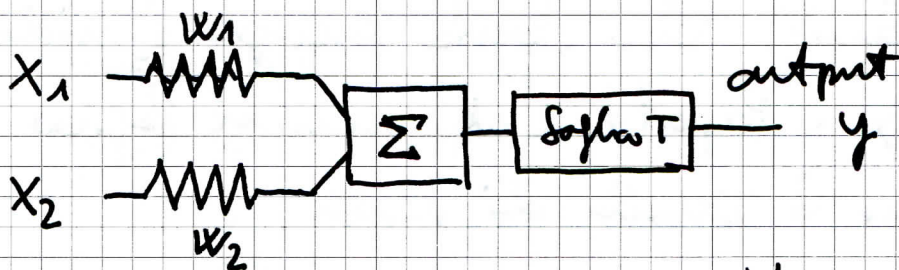


Valutare ora alcuni esempi di reti neurali

### Il modello di McCulloch e Pitts (MCP)

il modello venne sviluppato nel 1943 da McCulloch e Pitts e, come descritto nell'introduzione precedente, usava un modello di tipo elettrico, impostando i pesi delle connessioni come resistenze del circuito.

Analizziamo un semplice esempio in grado di realizzare la funzione logica AND



$X_1$	$X_2$	$y = X_1 \text{ AND } X_2$
0	0	0
0	1	0
1	0	0
1	1	1

L'esercizio richiede la ricerca dei pesi  $w_1$  e  $w_2$  tali per cui l'output soddisfi la funzione AND

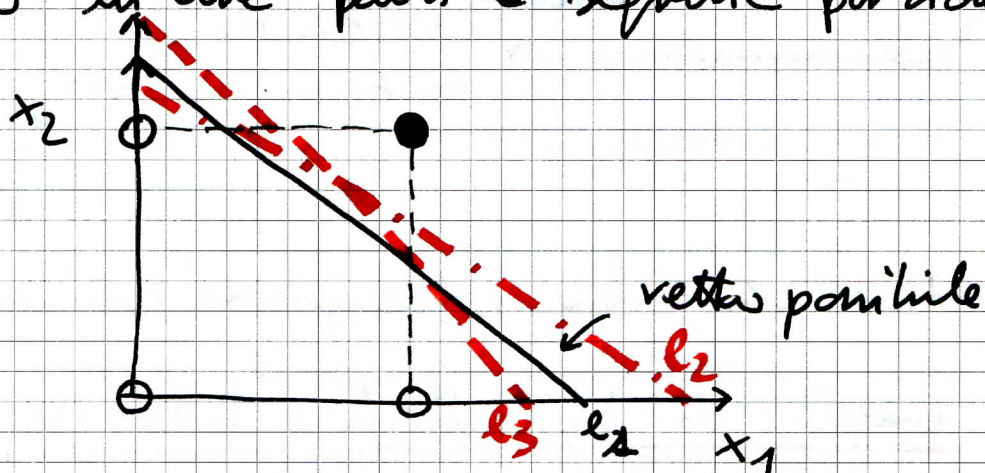


Si tratta di determinare la retta

$$w_1 x_1 + w_2 x_2 - T = 0$$

$$\Rightarrow x_2 = \frac{T}{w_2} - \frac{w_1}{w_2} x_1$$

Il che equivale a determinare la retta che separa in due parti le separate porzioni di piano



esistono infinite rette ( $l_1$ ,  $l_2$  e  $l_3$  sono esempi) che soddisfano le richieste.

Il problema è linearmente separabile ed è pertanto rappresentabile da una singola rete neurale (modello di McCulloch e Pitts).

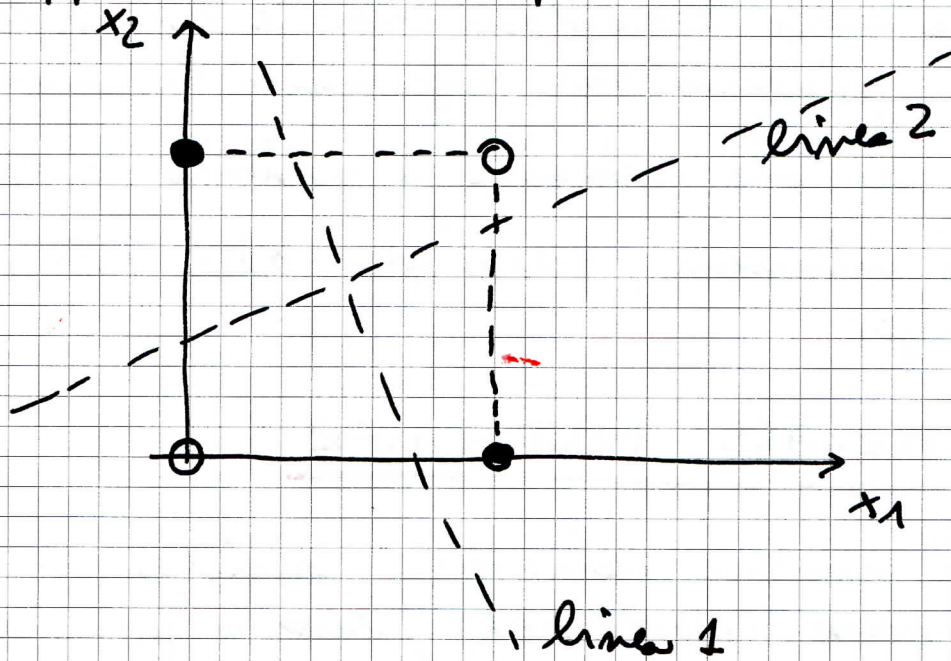
Purtroppo non tutti i problemi sono linearmente separabili. Un classico esempio (evidenziato da Minsky e Papert) è l'operazione logica XOR.

Vediamo la tabella di verità:

$x_1$	$x_2$	$y = x_1 \oplus x_2$
0	0	0
0	1	1
1	0	1
1	1	0



e rappresentando nel piano cartesiano

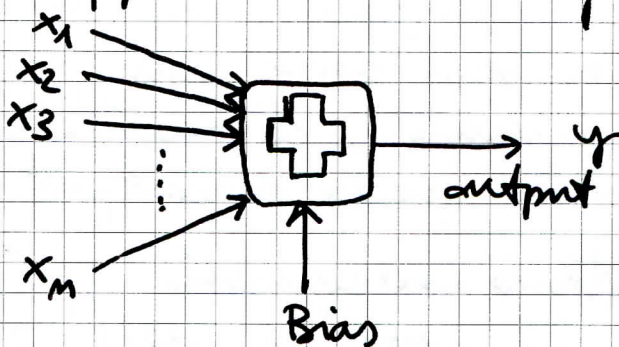


In queste due vettorie tra le due rette 'linea 1' e 'linea 2' è in grado di separare i due pallini pieni (●) da quelli vuoti (○).

## Il modello di Perceptron

È stato sviluppato da Frank Rosenblatt nel 1962.

È molto simile al MCP discusso in precedenza ed è rappresentato dal diagramma seguente:



La schema più semplice è costituito da una strato in ingresso (input layer) ed un neurone in uscita (output layer)

Sia i neuroni in ingresso che l'uscita possono assumere i valori +1 o -1.

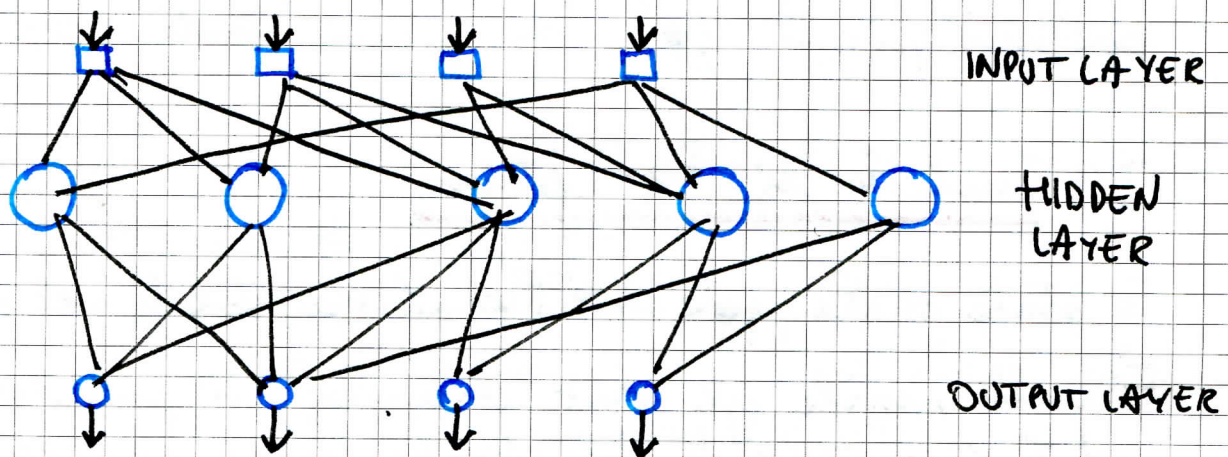


Nonostante i successi iniziali, la rete di tipo single layer Perceptron non è in grado di classificare problemi non separabili linearmente.

Per andare a tale problema è possibile introdurre degli strati intermedi, passando alle reti neurali multi-layer

## Reti FEEDFORWARD MULTI-LAYER

La rete più semplice realizzabile consiste di uno strato di ingresso (INPUT LAYER), uno o più strati intermedi di neuroni (HIDDEN LAYERS) e uno strato di output

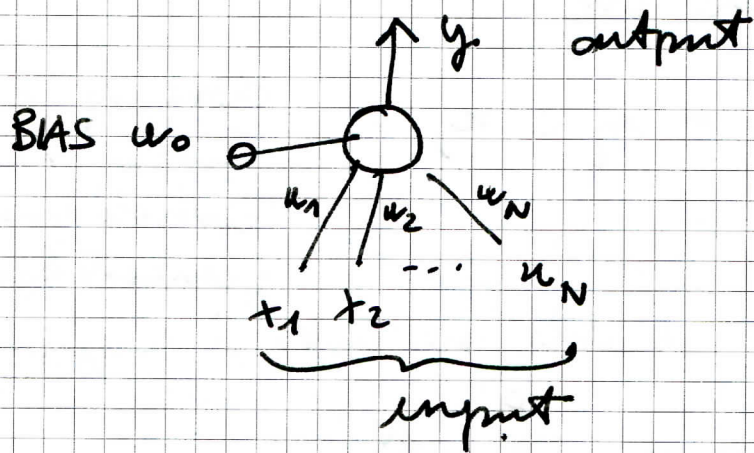


Le connessioni tra i neuroni sono uni-direzionali e sono tutte "in avanti", a partire dal layer di input fino a quello di output.

Una rete con uno strato intermedio è in grado di risolvere il problema dell'OR esclusivo.



Consideriamo ora un singolo neurone della rete



Se il neurone può assumere valori  $(0, 1)$ , la funzione di attivazione è

$$y(a) = \frac{1}{1 + e^{-a}} \quad \text{con } a = \sum w_i x_i$$

altrimenti, per un output tra  $(-1, 1)$

si sceglie

$$y(a) = \tanh(a)$$

Supponiamo di aver individuato un insieme di valori di input e di conoscere quale valore in output ci si aspetta: il campione prende il nome di campioni di training della rete

Definiamo una funzione e che tenga conto dell'errore commesso dalla rete nel riconoscimento

$$e^h = t^h - y^h$$

dove  $t^h \equiv$  output  
aspettato

$y^h \equiv$  output  
fornito dalla  
rete

ovvero  $h = 1, 2, \dots, P$

cioè come mi campioni di training.

Si vuole trovare una procedura che permetta di aggiustare i pesi  $w_i$ , in modo da ridurre gli



errori.

Si può dimostrare che

$$\Delta w_j = \eta e^h$$

dove  $\eta$  è un  
numero reale  
arbitrario  
(chiamato learning rate)

e l'errore è calcolato su tutti i campioni

$$e = \sum_{h=1}^P e^h$$

L'algoritmo prende il nome di "Gradient Descent"

- per ogni coppia di input/output  
 $\vec{x}^h, t^h$  con  $h=1, 2, \dots, P$

calcolare

$$y^h = y(\vec{x}^h, \vec{w})$$

- definire

$$e^h = t^h - y^h$$

e calcolare

$$g_j^h = -e^h w_j$$

- modificare i pesi in modo che

$$\Delta w_j = -\eta \sum_{h=1}^P g_j^h$$

La procedura permette di identificare i pesi migliori, date il campione di training selezionato.