

## **MODULO 3:**

# **CENNI SULLE RETI NEURALI**

---

Bibliografia:

D. Floreano, *Manuale sulle reti neurali*, Bologna, Il Mulino, 1996.

## 1. Sistemi di calcolo seriale e intelligenza naturale

Malgrado il fatto che i computer moderni siano sempre più potenti e veloci, è ancora molto difficile utilizzarli per risolvere alcuni problemi che per gli esseri umani sono relativamente banali. Il **riconoscimento di oggetti** in situazioni quotidiane, la **coordinazione motoria** necessaria per spostarsi da una stanza a un'altra e la valutazione contemporanea di un insieme di circostanze per poter prendere una rapida decisione sono tutti esempi di abilità che possediamo e che non comportano sforzi particolari. La programmazione di un computer per svolgere questi compiti si imbatte in difficoltà enormi. Una delle possibili cause sta nel fatto che il modo in cui i programmi tradizionali (basati su **regole e sistemi di conoscenze**) elaborano l'informazione è radicalmente diverso dal modo in cui funzionano i sistemi nervosi biologici.

Un computer tradizionale — anche detto **computer seriale** o computer di **von Neumann** — è composto sostanzialmente da un'unità di calcolo, o processore, che esegue in successione un altissimo numero di operazioni (nell'ordine di alcuni milioni al secondo) e da **tre tipi di memoria**:

- a) una che contiene le **istruzioni** necessarie a svolgere le operazioni;
- b) una **temporanea** da cui vengono letti i dati necessari e depositati i risultati dei calcoli effettuati;
- c) una **permanente** in cui questi dati rimangono registrati.

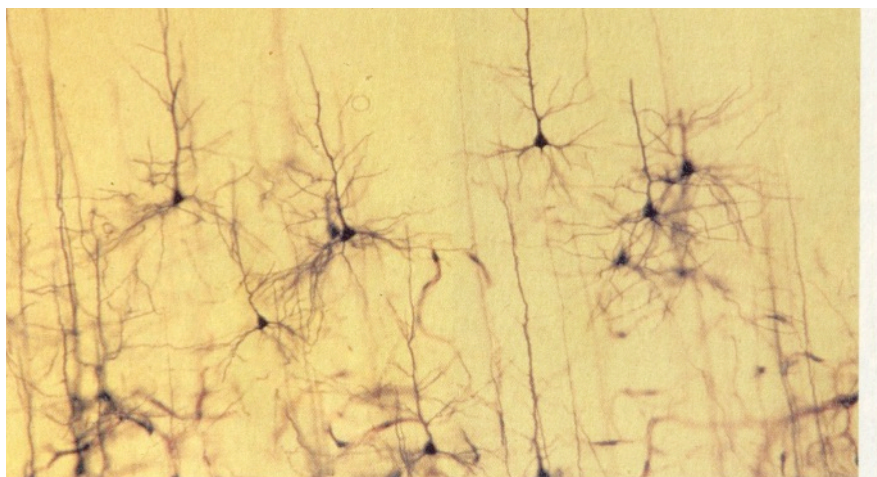
Per poter svolgere un qualsiasi compito i calcolatori tradizionali necessitano di un **programma** e la struttura dei programmi tradizionali riflette la struttura e il funzionamento dei computer su cui operano. Un programma tradizionale è composto di un insieme di istruzioni organizzate in modo gerarchico e da tabelle di consultazione ove vengono mantenute le conoscenze; in pratica il suo funzionamento consiste nel **leggere sequenzialmente** i dati e nell'applicare su di essi determinate operazioni in base alle regole e conoscenze predefinite.

**Il problema è che i sistemi di elaborazione seriale sono molto rapidi ed efficienti nel risolvere compiti in cui gli esseri umani trovano normalmente difficoltà (come, ad esempio, la soluzione di calcoli matematici, la rotazione di complicate figure geometriche, la memorizzazione esatta di enormi quantità di dati), ma si rivelano particolarmente inefficienti e lenti nell'affrontare compiti che appartengono alla nostra vita quotidiana.**

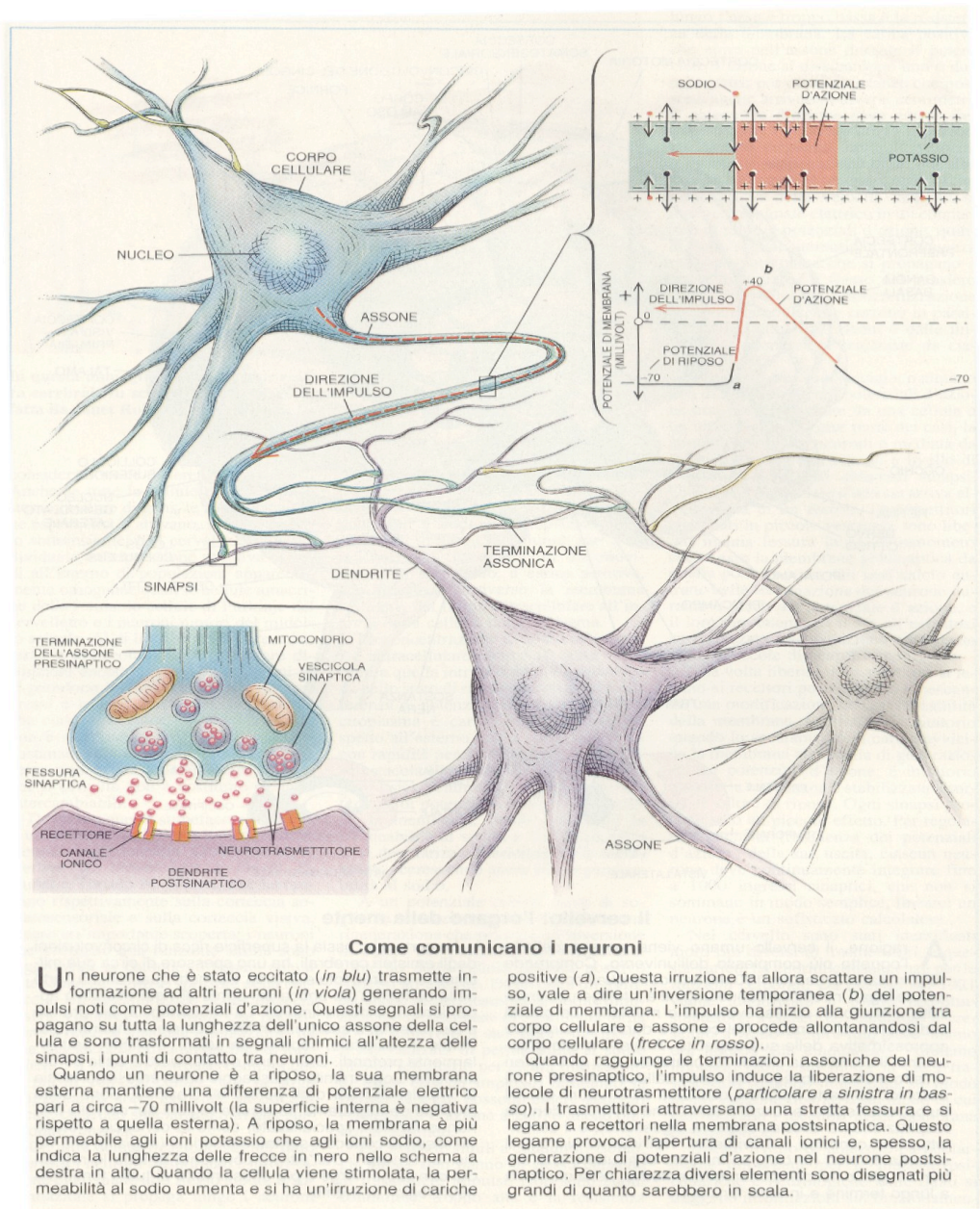
Se osserviamo in maggior dettaglio la natura di questi due tipi di compiti, notiamo che i primi — quelli propri del mondo dei calcolatori — sono descrivibili da una serie di regole o procedure e possiedono una **soluzione analitica**, mentre i secondi — quelli tipici della vita quotidiana degli esseri umani — sono **difficili da descrivere attraverso regole esplicite**, non sempre è possibile ricavarne una soluzione analitica.

Questa differenza di rendimento trova riflesso anche nella struttura dei due sistemi di elaborazione sottostanti.

Al contrario del calcolatore seriale, il **sistema nervoso centrale** (il cervello) umano contiene circa  $10^{11}$  elementi di elaborazione (**neuroni**) ciascuno dei quali **comunica in media con altri  $10^4$  elementi**. Malgrado alcune differenze fisiologiche, è ragionevole ipotizzare che i neuroni funzionino pressappoco in modo simile: **ciascuno di essi emette una risposta in funzione del segnale globale ricevuto e della propria soglia di attivazione**.



In questa immagine i neuroni sono evidenziati con il metodo di Golgi. L'architettura cerebrale fu scoperta da Santiago Ramón y Cajal. Questa preparazione è stata fatta da Janet Robbins del gruppo di David H. Hubel alla Harvard Medical School.



Non esiste una porzione specifica del sistema nervoso che si occupi solamente di immagazzinare le conoscenze e un'altra che applichi esclusivamente delle regole; inoltre il sistema di trasmissione dei segnali — al contrario di quanto avviene nei computer tradizionali — è molto **rumoroso** (ovvero la decisione di emettere un segnale è meglio descritta dalla *probabilità* di emettere il segnale) e non molto affidabile (molti elementi e linee di comunicazione tendono a morire in modo abbastanza casuale sia per processi di normale invecchiamento che per motivi patologici).

**Il funzionamento del sistema nervoso è radicalmente diverso dal funzionamento di un sistema di elaborazione seriale dell'informazione.**

Le differenze principali riguardano i seguenti aspetti:

➤ l'elaborazione dell'informazione nei sistemi nervosi avviene in *parallelo* mentre nei calcolatori tradizionali ciascun dato viene elaborato individualmente e in successione. Malgrado il fatto che ogni singolo neurone sia relativamente lento [Un neurone è in grado di emettere qualche *centinaio* di impulsi per secondo, mentre l'unità di calcolo di un computer di media potenza elabora parecchi *milioni* di numeri interi per secondo].

Il parallelismo massivo spiega — in parte — la maggior velocità del cervello nell'eseguire compiti che richiedono l'elaborazione contemporanea di un elevato numero di dati, come ad esempio il riconoscimento visivo di oggetti.

- L'elaborazione nei sistemi nervosi è *distribuita* su molti elementi, ovvero vi sono molti neuroni che si occupano della stessa operazione. L'osservazione di un sistema nervoso — attraverso le tecniche di registrazione intracellulare o di visualizzazione dell'attività cerebrale (*PET* o *MRI*) — durante lo svolgimento di semplici compiti evidenzia l'attivazione contemporanea di molti neuroni, a volte organizzati in gruppetti locali, altre volte distribuiti «a macchie» in zone diverse del cervello. Inoltre un singolo neurone può prender parte in diversi tipi di operazioni eseguibili sia contemporaneamente che in tempi diversi.
- Ogni dato nella memoria dei calcolatori è identificato da un *indirizzo* (in pratica un numero) che viene utilizzato dal processore centrale per recuperare le conoscenze necessarie allo svolgimento di un certo compito. Invece gli esseri umani accedono alle proprie memorie in base al *contenuto*: noi siamo in grado di recuperare un ricordo semplicemente in base a qualche indizio parziale.
- **I sistemi nervosi, al contrario dei calcolatori, non devono essere programmati per svolgere un compito, bensì *imparano* autonomamente in base all'esperienza o con l'aiuto di un insegnante esterno.**



**Un calcolatore necessita invece di un programma che contiene tutte le istruzioni necessarie per portare a termine il compito correttamente.**

In conclusione i **computer seriali** e i relativi programmi tradizionali sono degli strumenti molto potenti per svolgere dei compiti che richiedono la ripetizione di una serie di operazioni ben definite ove l'accuratezza, l'**affidabilità** e la **velocità** sono le caratteristiche importanti.

## 2. Che cos'è una rete neurale artificiale?

Le reti neurali artificiali sono dei sistemi di elaborazione dell'informazione il cui funzionamento trae ispirazione dai sistemi nervosi biologici. Una rete **neurale artificiale possiede molte semplici unità di elaborazione variamente connesse tra di loro**. Alcune di queste unità ricevono informazioni dall'ambiente (*input*), altre emettono risposte nell'ambiente (*output*) e altre ancora comunicano solamente con le unità all'interno (*hidden*). **Ciascuna unità intende simulare il ruolo di un neurone** — o di un gruppo di neuroni — nelle reti neurali biologiche: per questo motivo esse vengono anche definite impropriamente *neuroni*. Altri nomi comunemente utilizzati sono *nodo* e *processore*. **Ciascuna unità svolge un'operazione molto semplice che consiste nel diventare attiva se la quantità totale di segnale che riceve supera la propria soglia di attivazione.**

Se un'unità diventa attiva, essa emette un segnale che viene trasmesso lungo i canali di comunicazione fino alle altre unità a cui essa è connessa; ciascun punto di connessione agisce come un filtro che trasforma il messaggio ricevuto in un segnale *inibitorio* o *eccitatorio* aumentandone o diminuendone nel contempo l'intensità a seconda delle proprie caratteristiche individuali.

Questi punti di connessione simulano le *sinapsi* biologiche da cui prendono spesso anche il nome; inoltre, poiché il loro ruolo consiste in effetti nel «pesare» l'intensità dei segnali trasmessi, essi vengono definiti anche con il nome di *pesi sinaptici* o semplicemente *pesi*.

Quindi: un neurone artificiale è caratterizzato da un insieme di **sinapsi** che corrispondono ai terminali di altri neuroni, da una **soglia** e da una **funzione di attivazione**; esso verrà indicato anche con il nome di **unità, nodo** e processore.

L'effetto di un segnale  $x$  sul **neurone postsinaptico** è semplicemente uguale al prodotto

$$w x,$$

dove  $w$  è la sinapsi [peso] corrispondente.

L'*input netto*  $A_i$ , di un neurone  $i$ -esimo è la la somma algebrica dei prodotti fra tutti i  $j$  segnali di ingresso  $x_j$  e i valori delle sinapsi corrispondenti  $w_{ij}$ :

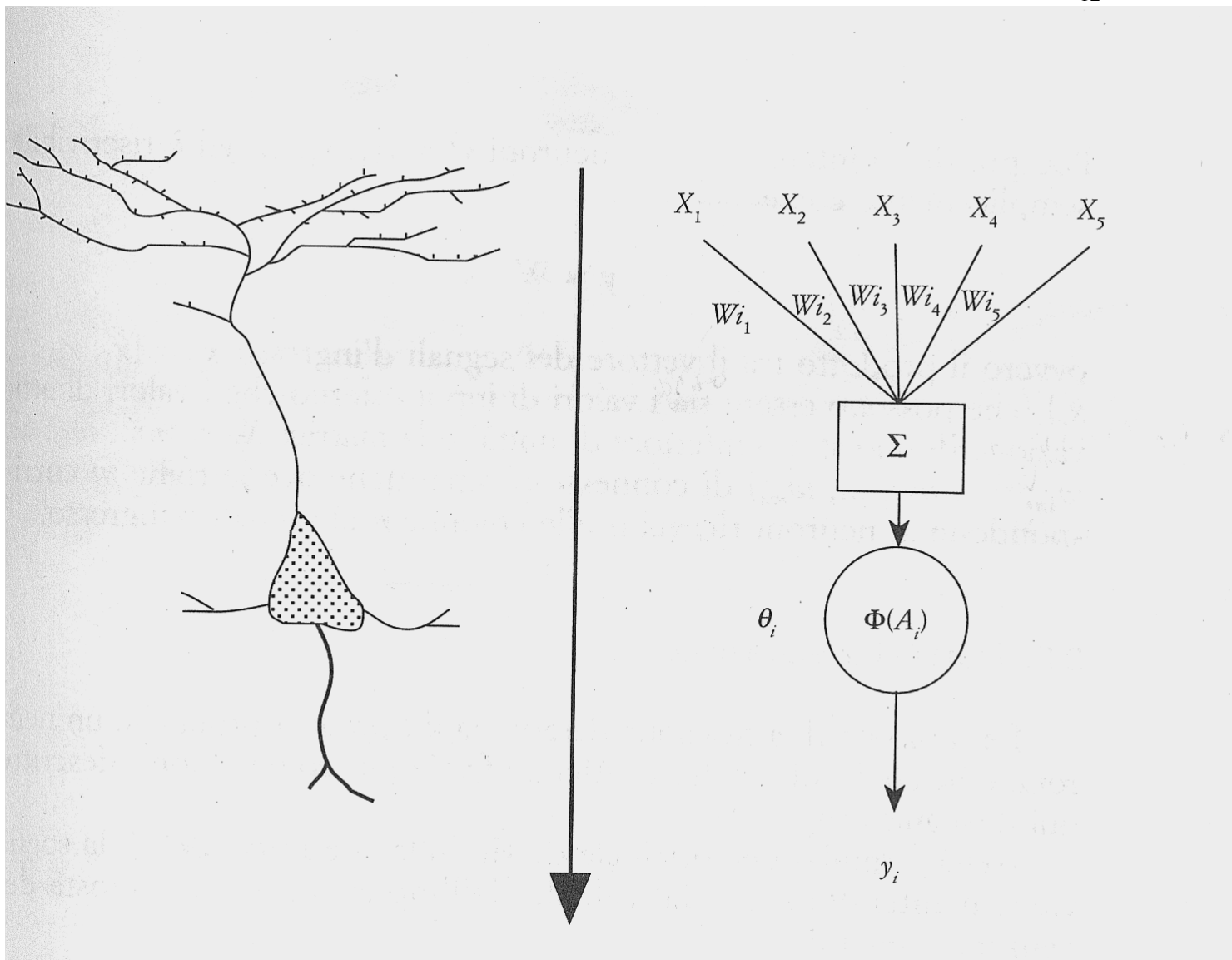
$$\sum_j^N w_{ij} x_j$$

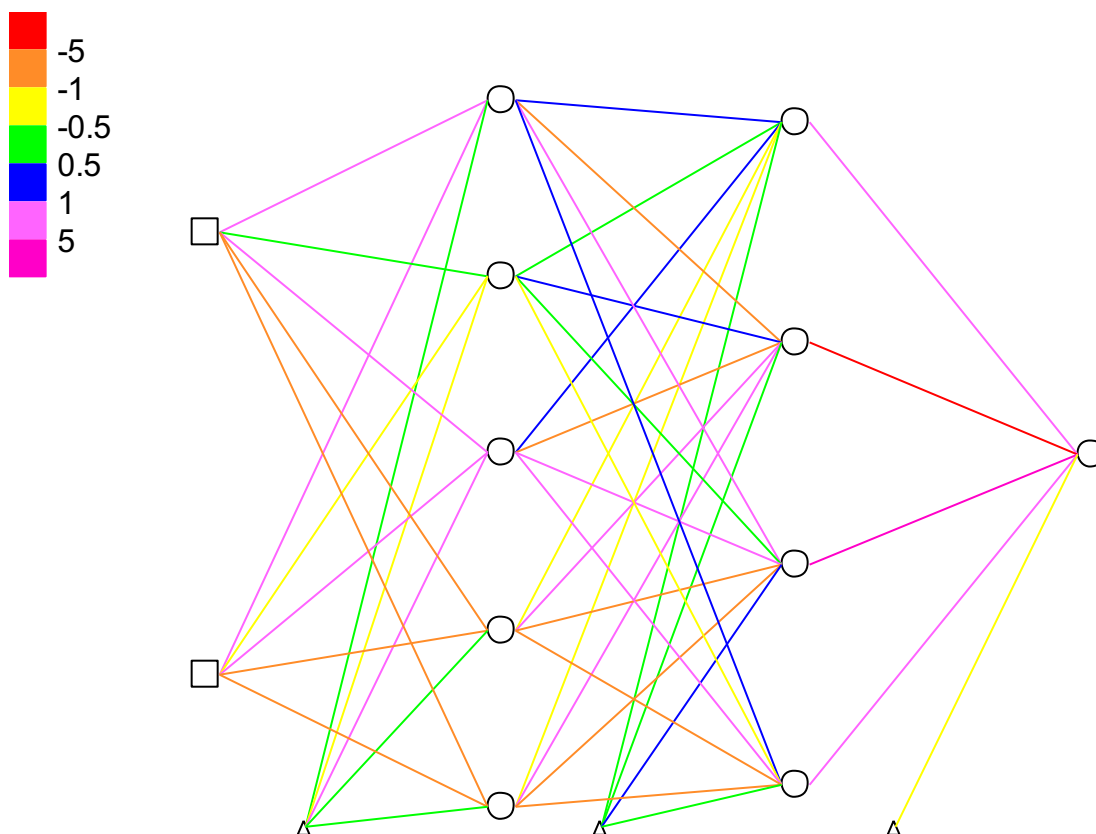
a cui si sottrae il **valore di soglia**  $\vartheta_i$  del neurone:

$$A_i = \sum_j^N w_{ij} x_j - \vartheta_i$$

La risposta del neurone  $y_i$  è ottenuta passando l'input netto attraverso una **funzione d'attivazione**  $\Phi(x)$ :

$$y_i = \Phi(A_i) = \Phi\left(\sum_j^N w_{ij} x_j - \vartheta_i\right)$$

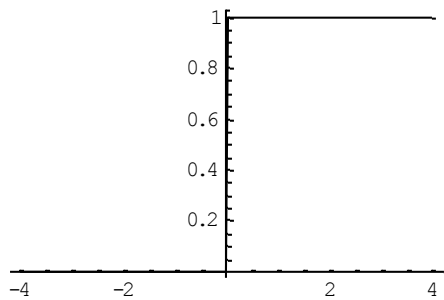




**Rete neurale a 2 livelli (*layer*) con 2 unità input, 1 unità output e 9 neuroni hidden**

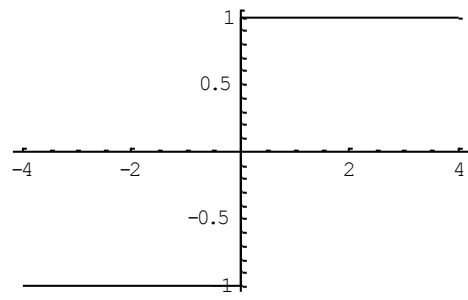
## Alcune delle più comuni **funzioni di attivazione**:

[Le prime due funzioni vengono qui definiti sull'input e non sull'input netto]  
 $x := \text{input}$ ;  $\vartheta := \text{soglia}$



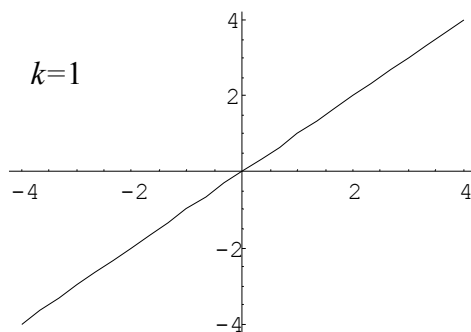
Funzione a **gradino**

$$\Phi(x) = \begin{cases} 1, & \text{se } x > \vartheta; \\ 0, & \text{altrimenti.} \end{cases}$$



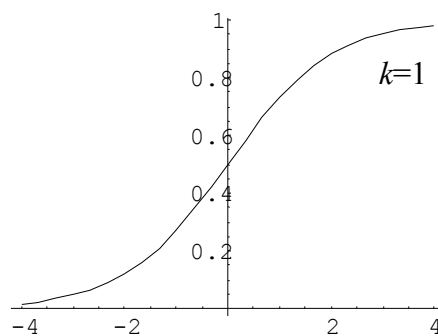
Funzione **bipolare**

$$\Phi(x) = \begin{cases} 1, & \text{se } x > \vartheta; \\ -1, & \text{altrimenti.} \end{cases}$$



Funzione **lineare**

$$\Phi(x) = kx, \quad k \in \mathfrak{R}$$



Funzione **logistica (sigmoide)**

$$\Phi(x) = \frac{1}{1 + e^{-kx}}, \quad k \in \mathfrak{R}$$

## Come avviene l'apprendimento della rete neurale?

Uno dei metodi più usati è l'*apprendimento supervisionato*.

La modifica dei valori sinaptici avviene impiegando **una misura di errore tra la risposta fornita dalla rete neurale e la risposta desiderata per ogni input**.

### 1. Training

I valori iniziali dei pesi sinaptici della rete vengono assegnati in modo casuale entro un piccolo campo di variazione. L'apprendimento consiste nella presentazione ripetuta di una coppia, costituita dall'input e dalla risposta desiderata. Mediante algoritmi di apprendimento [per esempio *backpropagation*], vengono modificati i pesi sinaptici.

### 2. Test

Una volta terminata la fase di apprendimento, i pesi sinaptici vengono "congelati" ed è possibile studiare la risposta della rete su nuovi input.

La capacità di apprendere rappresenta forse l'elemento di maggiore attrazione dei modelli neurali perché permette di impiegare una rete neurale per risolvere problemi **senza dover individuare direttamente la soluzione analitica**, ma semplicemente esponendo il modello neurale a una serie di esempi.

### 3. Impiego di reti neurali artificiali

Le reti neurali artificiali presentano alcune caratteristiche che si rivelano interessanti in molti campi di ricerca e domini di applicazione. Benché molte di queste caratteristiche varino da modello a modello, ve ne sono alcune sufficientemente generali:

**Robustezza:** una rete neurale è resistente al rumore, ovvero è in grado di continuare a dare una risposta corretta anche se alcune delle sue connessioni vengono eliminate ("lesionate") o se viene aggiunto del rumore al segnale d'ingresso, ai canali di trasmissione o alla funzione di attivazione dei nodi. Questa proprietà è comune anche ai sistemi nervosi biologici ove la capacità di apprendere e ricordare non viene alterata in modo sostanziale dalla perdita continua di neuroni. A mano a mano che il livello di rumore aumenta le prestazioni delle reti neurali artificiali subiscono un «decadimento graduale»: la percentuale di errore potrebbe aumentare in modo lento e pressappoco uniforme su tutto il campo di risposta oppure la rete potrebbe perdere la capacità di rispondere correttamente ad alcuni stimoli e mantenere una risposta inalterata per tutti gli altri. Inoltre, come nel caso dei sistemi nervosi biologici, le reti neurali artificiali lesionate possono essere talvolta riaddestrate ad acquistare le abilità perse. Queste proprietà rappresentano un vantaggio rispetto alle modalità di funzionamento dei sistemi seriali ove di solito la perdita di un singolo anello della catena di



elaborazione comporta una caduta catastrofica delle prestazioni dell'intero sistema.

***Flessibilità:*** un modello neurale può essere impiegato per un grande numero di finalità diverse: esso non ha bisogno di conoscere le proprietà del dominio specifico di applicazione perché le apprende in base all'esperienza. Questo non significa che un qualsiasi modello neurale possa essere utilizzato per tutti i tipi di compiti, ma implica che l'utente non deve necessariamente conoscere le soluzioni dettagliate e analitiche che caratterizzano il problema sotto indagine. In generale l'utente di una rete neurale deve essere in grado di individuare precisamente le finalità del progetto, il tipo di compito e una serie di vincoli al fine di valutare qual è il modello neurale che risulta più appropriato. Ulteriori conoscenze sulla natura del problema possono essere utilizzate per decidere altri aspetti importanti del modello neurale come ad esempio l'architettura, il tipo di codifica degli stimoli esterni e della risposta della rete, alcuni parametri di addestramento, ecc. Da un lato questa caratteristica presenta notevoli vantaggi perché permette di affrontare molti problemi di cui non sono note le soluzioni analitiche; dall'altro lato vi è però il pericolo di rinunciare a cercare di comprendere a fondo la natura di un problema e di rifugiarsi in una «soluzione neurale» che non aumenta la nostra conoscenza.

**Generalizzazione:** una rete neurale che è stata addestrata su un numero limitato di esempi è in grado di produrre una risposta adeguata a *input* d'ingresso che non ha mai visto in precedenza, ma che presentano tuttavia qualche somiglianza con gli esempi presentati durante la fase di addestramento. La capacità di generalizzare a nuovi stimoli è una caratteristica molto apprezzata nei tipici campi di applicazione delle reti neurali ove spesso è impossibile ottenere una collezione esaustiva di tutti i dati su cui la rete neurale dovrà operare.

**Recupero in base al contenuto:** le reti neurali artificiali sono in grado di recuperare le proprie memorie in base al contenuto partendo da dati incompleti, simili o corrotti da rumore. Come nei sistemi nervosi biologici, è sufficiente un indizio per dirigere l'attivazione del sistema nella direzione appropriata completando e recuperando l'intera memoria. In alcuni modelli il processo di recupero presenta modalità simili a quelle esibite dai soggetti umani: gli oggetti familiari vengono riconosciuti più rapidamente di quelli sconosciuti. I calcolatori seriali invece recuperano i dati impiegando un numero che rappresenta l'indirizzo di memoria corrispondente; se questo numero viene alterato o perso, non è più possibile recuperare l'intero dato.

Le reti neurali artificiali vengono impiegate in molti settori di ricerca e campi di applicazione.

### **Informatica:**

- *Compressione di dati*
- *Eliminazione del rumore*
- *Riconoscimento di segnali sonar*
- *Riconoscimento caratteri*
- *Riconoscimento vocale*

### **Analisi finanziaria**

### **Medicina**

### **Psicologia:**

- *Studio dei processi cognitivi*