

Capitolo undicesimo

Apprendimento e memoria nelle reti neurali

Dai neuroni al comportamento

Le reti neurali artificiali sono sistemi di elaborazione dell'informazione che

- traggono ispirazione dal funzionamento del cervello
- sono in grado di *apprendere* a svolgere compiti complessi

Esse hanno numerose applicazioni tecnologiche:

- riconoscere immagini o suoni
- controllare i movimenti di un robot
- predire fenomeni complessi (per es. andamento dei mercati azionari)

Le reti neurali artificiali non si propongono di riprodurre il neurone biologico in tutta la sua complessità

Esse invece cercano di riprodurre i *principi base di funzionamento* dei sistemi biologici

Nella psicologia e nelle neuroscienze le reti neurali non sono utilizzate per riprodurre le funzioni mentali *in chiave riduzionistica* (= ridurre la psicologia alla neurofisiologia o alla biologia molecolare)

Esse invece «ricostruiscono» tali funzioni come *proprietà emergenti* dei sistemi neuronali

Reti neurali: elementi di base

Reti neurali biologiche

Le reti neurali biologiche sono formate da neuroni che si influenzano reciprocamente attraverso le connessioni (*sinapsi*) che li collegano

- Ogni neurone *rileva* un certo insieme di condizioni e *segnala* ciò che ha rilevato attraverso la sua frequenza di scarica
- Attraverso i segnali che si scambiano i neuroni possono formare *sistemi di rilevatori* più complessi

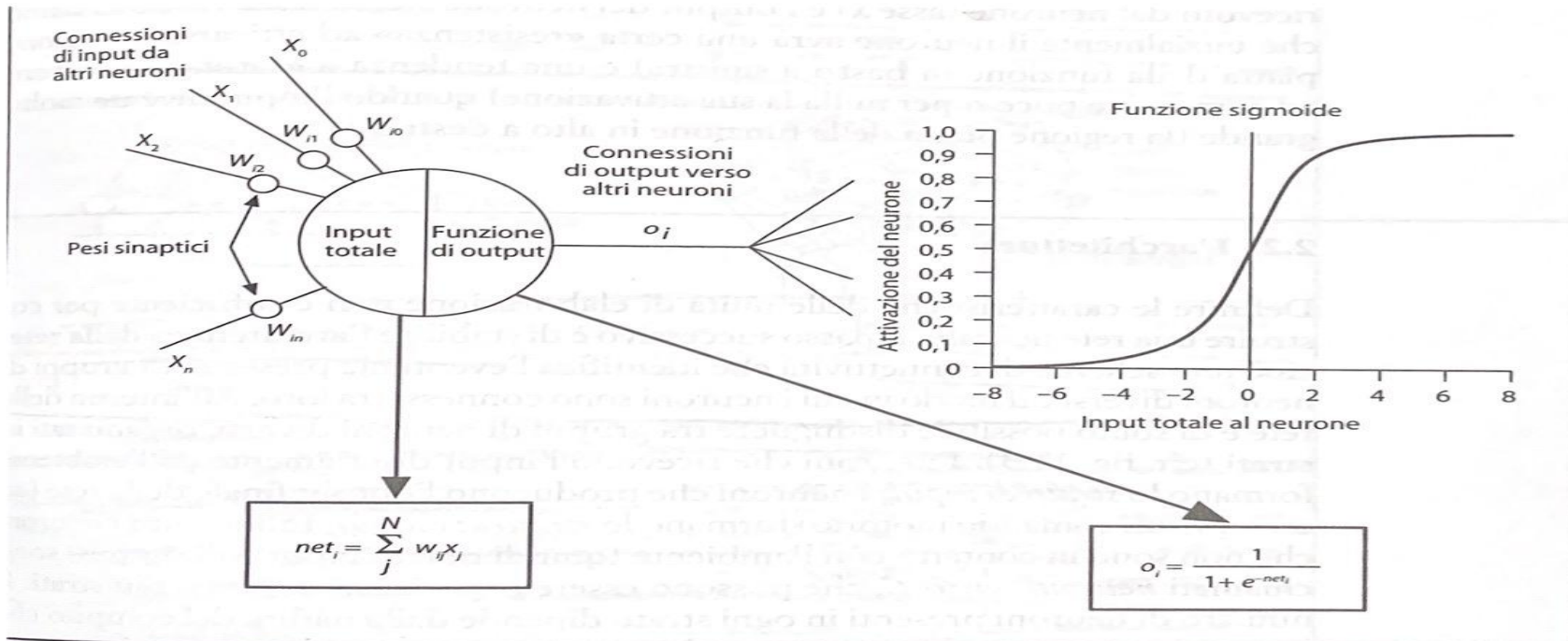
Il neurone formale

Un *neurone formale* è un modello matematico che cerca di catturare gli aspetti fondamentali del funzionamento neuronale

Un neurone è caratterizzato da uno *stato di attivazione*, che varia in funzione delle attivazioni dei neuroni dai quali riceve e che viene propagato ai neuroni ai quali trasmette

Ogni connessione tra neuroni ha un *peso* che indica la forza della connessione e può essere *positivo* (eccitatorio) o *negativo* (inibitorio)

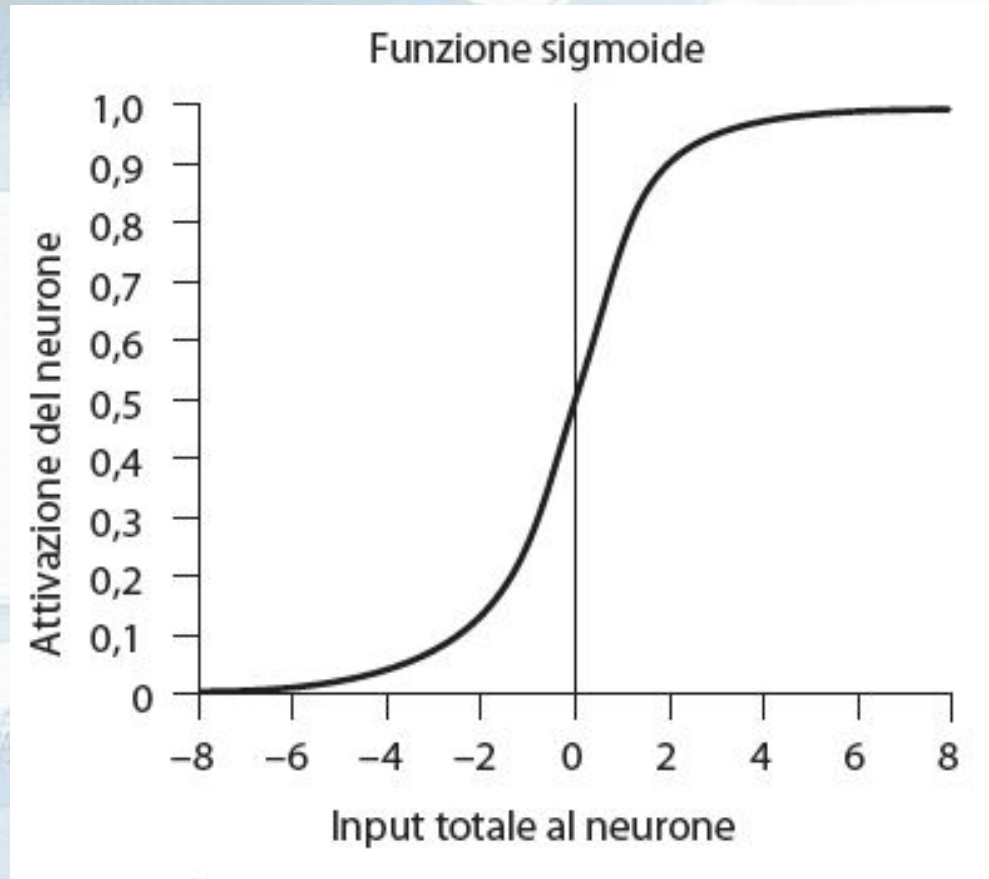
L'*input* che un neurone riceve da ciascuno dei neuroni cui è connesso si calcola *moltiplicando* il *segnale* proveniente da quel neurone per il *peso* sulla connessione



L'*input totale* del neurone è la *sommatoria* delle attivazioni che il neurone riceve dai neuroni vicini

Lo stato di attivazione del neurone, che viene inviato ai neuroni «a valle», viene calcolato a partire dall'input totale attraverso la *funzione di output*

La tipica funzione di output è una funzione *non-lineare a forma di S*



Questa funzione esprime la relazione tra l'input totale del neurone (asse x) e l'output del neurone stesso (asse y)

Il neurone mostra una *resistenza* iniziale ad attivarsi (la regione piatta della funzione in basso a sinistra) e una tendenza a *saturarsi* quando l'input diventa molto grande (la regione piatta della funzione in alto a destra)

Architettura delle reti neurali

Una rete è caratterizzata da una architettura che identifica

- gruppi o *strati* di neuroni diversi che costituiscono le rete
- il modo in cui i neuroni sono *connessi* tra di loro

Strati di unità

le unità che ricevono l'input direttamente dall'ambiente formano lo *strato di input*

le unità che producono l'output finale della rete formano lo *strato di output*

le unità che non sono in contatto diretto con l'esterno sono chiamate *unità nascoste*

Schemi di connettività

Associatore di configurazioni: sono presenti solo unità di input e unità di output, le connessioni sono unidirezionali

Rete feed-forward: sono presenti connessioni unidirezionali dalle unità di input alle unità nascoste e da queste alle unità di output

Rete ricorrente: sono presenti connessioni bidirezionali, in cui l'attivazione può fluire anche all'indietro dall'output verso l'input

Rete ricorrente con connessioni laterali: come la rete ricorrente, ma sono presenti anche connessioni intra-strato

Apprendimento delle reti neurali

L'*apprendimento* in una rete consiste nel trovare l'insieme di *pesi sulle connessioni* che permette alla rete di produrre l'output desiderato dato un certo input

Regola di Hebb: se due neuroni collegati fra loro sono *contemporaneamente attivi*, l'efficacia sinaptica (= peso sulla connessione) *aumenta*

Tipi di algoritmi di apprendimento

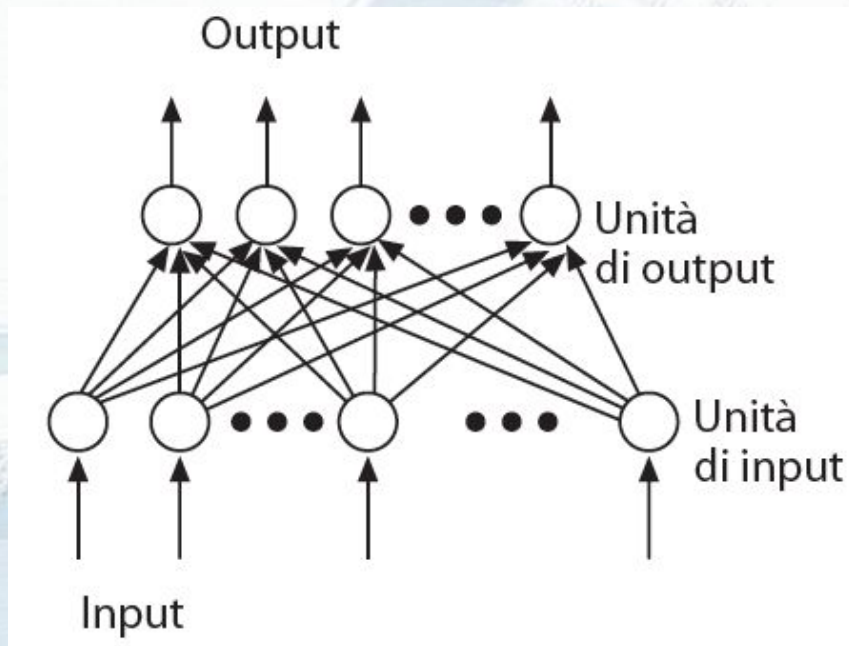
Apprendimento *supervisionato*: alla rete viene fornito l'output desiderato, come se fosse presente un insegnante esterno

Apprendimento *per rinforzo*: alla rete vengono fornite informazioni sulla bontà dell'output prodotto; l'apprendimento ha lo scopo di imparare ad agire in modo da massimizzare il rinforzo nel lungo termine

Apprendimento per *auto-organizzazione*: la rete costruisce rappresentazioni dell'input sempre più complesse e informative, non vi è un output desiderato

Associatore di configurazioni

Struttura di un semplice *associatore di configurazioni*



Il compito di una rete siffatta è associare determinate configurazioni di input a determinate configurazioni di output

Una semplice procedura di apprendimento supervisionato per associare configurazioni è la *regola delta*

Regola delta

- viene presentata alla rete una configurazione di input
- l'attivazione fluisce alle unità di output
- viene calcolata l'attivazione delle unità di output
- la configurazione così ottenuta viene confrontata con la configurazione di output desiderata
- viene calcolata la *discrepanza* tra le due configurazioni (= segnale di errore)
- i pesi sulle connessioni vengono modificati *in modo da ridurre l'errore*
- l'apprendimento termina quando l'errore arriva a 0 (o è sufficientemente piccolo)

Dopo l'apprendimento la rete può essere in grado di *generalizzare*, cioè di rispondere correttamente anche di fronte a nuovi esempi del problema di apprendimento

Reti neurali e memoria

Si può dire che le reti memorizzano ciò che apprendono?

Nelle reti neurali vi sono due tipi di memoria

- i *pesi sulle connessioni* rappresentano conoscenze a lungo termine
- l'*attivazione delle unità* rappresenta le informazioni specifiche possedute dalla rete in un determinato istante di tempo

Interferenza

Le reti neurali mostrano anche alcuni fenomeni relativi alla memoria simili a quelli osservabili negli esseri umani

Gli studi sulla memoria umana hanno portato alla luce il fenomeno dell'*interferenza*

È stato dimostrato sperimentalmente che l'apprendimento di materiale *nuovo* può interferire con quanto *già appreso*

Compito di apprendimento associativo AB-AC

Vengono presentate ai partecipanti due liste di coppie di parole

- *lista AB*: ogni coppia della lista è costituita da una parola del gruppo A e da una parola del gruppo B
- *lista AC*: ogni coppia della lista è costituita da una parola del gruppo A e da una parola del gruppo C

I partecipanti dopo aver studiato la lista AB rievocano la parola B appropriata per ogni parola A

Poi studiano la lista AC e vengono interrogati su entrambe le liste

I risultati mostrano una *interferenza delle associazioni apprese successivamente (AC) su quelle apprese per prime (AB)*

Interferenza e integrazione

Se una rete neurale viene sottoposta al compito di apprendimento associativo si ottiene un effetto di interferenza ancora maggiore che negli esseri umani

Il principale fattore alla base dell'interferenza è il *grado di sovrapposizione* delle rappresentazioni

Ma d'altra parte la sovrapposizione delle rappresentazioni è utile perché permette l'*integrazione* e la *generalizzazione* delle conoscenze

Il cervello umano ha risolto il problema del contrasto fra interferenza e integrazione evolvendo due sistemi di apprendimento *separati e complementari*:

- *ippocampo*: specializzato nell'apprendimento rapido e non soggetto a interferenza
- *neocorteccia*: apprende lentamente e integra gradualmente le esperienze estraendo le conoscenze generali sul mondo