Slide 1: Titolo

Slide 2: Cosa è una rete neurale

Slide 3: Storia con McCulloch, Pitts e Rosenblatt

Slide 4: Applicazioni

Slide 5: Sistema nervoso umano

Slide 6: Modello formale di un neurone artificiale

Slide 7: Attivazione e varie funzioni di attivazione

Slide 8: Architettura ( reti ff, totalmente connesse, ricorrenti)

Slide 9: Addestramento

Slide 10: Paradigmi (superv e non superv)

Slide 11: Processi

Slide 12: Attività di apprendimento

Slide 13: Cosa sono le reti di Hopfield

Slide 14: Architettura ed aggiornamento

Slide 15: Stabilità e funzione di energia

Slide 16: Caso discreto

**Slide 6: Modello formale di un neurone artificiale**

Questa breve illustrazione del sistema neurale umano suggerisce uno schema per delineare l’organizzazione delle reti neurali artiﬁciali. Queste ultime sono formate da un elevato numero di unità computazionali, che possono essere equiparate ai neuroni umani, capaci di eseguire una somma pesata. Tra gli elementi base che formano queste unità troviamo:

• un insieme di connessioni;

• un sommatore;

• una funzione di attivazione;

• un valore di soglia

Le unità sono collegate tra loro attraverso delle connessioni, così come le sinapsi collegano i neuroni nella rete umana. Consideriamo una generica unità costituita da canali di ingresso . Gli input provenienti da strati precedenti o dall’esterno entrano nel neurone tramite tali canali. Sulle connessioni sono presenti dei pesi sinaptici wi, numeri reali che denotano l’eﬃcacia sinaptica, ovvero la forza della connessione. Se wi >0 il canale è detto eccitatorio, se wi<0 il canale è inibitorio. I segnali in entrata sono convogliati nel soma del neurone artiﬁciale all’interno del quale vengono sommati producendo una combinazione lineare indicata con la parola net ed il segnale con cui il neurone trasmette la sua attività all’esterno è calcolato applicando una funzione di attivazione ϕ che limita l’ampiezza dell’output. Il modello neuronale include anche un valore soglia che ha l’eﬀetto, a seconda della sua positività o negatività, di aumentare o diminuire il valore in ingresso alla funzione di attivazione. Interpretando la soglia come il peso associato ad un ulteriore canale di ingresso x0, e quindi w0 = θ, l’output ﬁnale sarà:

**Slide 7: Attivazione e funzioni di attivazione**

Nelle reti neurali si può avere una attivazione asincrona in cui i neuroni si attivano uno per volta ed un’attivazione sincrona in cui tutti i neuroni si attivano contemporaneamente.

Gli step per l’attivazione di un neurone sono:

1. i valori di input arrivano al neurone e vengono combinati con i relativi pesi;

2. i prodotti risultanti vengono sommati e tale somma viene confrontata con un valore di soglia che dipende dalla funzione di attivazione del neurone;

3. se la somma supera il valore di soglia, il neurone si attiva inviando output sulle connessioni pesate in uscita, altrimenti viene inibito.

La funzione di attivazione determina il tipo di risposta che un neurone è in grado di emettere. Deﬁnisce, quindi, l’uscita di un neurone in funzione del livello di attivazione. L’uscita può essere un numero reale, un numero reale appartenente ad un intervallo, oppure un numero appartenente ad un insieme discreto.

Esistono varie tipologie di funzioni di attivazione, le principali sono: funzioni a soglia, funzioni lineari, funzioni lineari a tratti, funzione sigmoide.

* Iniziamo con la funzione a soglia che è tra le più utilizzate. Imponendo

il valore di uscita è ed assume questa forma (figura)

* Per quanto riguarda la funzione lineare e la funzione lineare a trattiimponendo sempre otteniamo come valore di uscita y = ϕ(a) = a per quanto riguarda la lineare mentre un esempio di output in presenza di un lineare a tratti è

y = ϕ(a)=

e rappresentandole si hanno questi grafici

* Infine per quanto concerne la funzione sigmoide, questa risulta essere appartenente alla famiglia di funzioni continue non lineari ed insieme alla funzione a soglia risulta essere tra le più utilizzate. In particolare riportiamo una funzione appartenente a questa classe che prende il nome di funzione logistica ed è definita secondo questa forma y = ϕ(a) =

mentre la sua rappresentazione è quella in figura (figura)

**Slide 8: Architettura (reti ff, totalmente connesse, ricorrenti)**

L’architettura di una rete neurale artiﬁciale è caratterizzata da:

• numero di strati di sinapsi;

• numero di neuroni presenti nell’input layer ovvero lo strato di ingresso;

• numero di neuroni nell’output layer, lo strato di uscita.

Le reti neurali si suddividono principalmente in tre grandi classi: le reti completamente connesse, le reti feedforward e le reti feedback o ricorrenti.

* Reti completamente connesse:

Nelle reti completamente connesse ogni neurone è connesso con tutti gli altri. Le connessioni tra i neuroni sono bidirezionali e possono essere rappresentate per mezzo di una matrice quadrata W di dimensione pari al numero di neuroni. Un suo generico elemento wi,j rappresenta il peso della connessione tra il neurone i ed il neurone j.

* Reti feedforward

In queste reti i segnali viaggiano dallo strato di ingresso verso lo strato di uscita e non esistono connessioni tra i neuroni all’interno di uno stesso strato, nè tra neuroni di strati non limitroﬁ. Ogni neurone in un generico strato è connesso con tutti quelli dello strato successivo ed i neuroni dello strato di ingresso hanno come unico compito quello di trasmettere i segnali ricevuti allo strato successivo, all’interno di essi non avviene alcuna computazione. Le reti feedforward a strati si distinguono in base al numero di strati che presentano, numero che dipende dallo speciﬁco problema che si intende risolvere.

- Reti feedforward ad uno strato che è la forma più semplice.

- Reti feedforward a più strati dette anche reti multilivello , in inglese Multi-Layer Perceptron, MLP. Tra l’input layer e l’output layer sono presenti uno o piu` strati di neuroni nascosti, si parla quindi di hidden layers. Nelle MLP ogni strato ha connessioni entranti dal precedente strato e uscenti in quello successivo, quindi la propagazione del segnale avviene in avanti in modo aciclico e senza connessioni trasversali.

- Reti ricorrenti che si distinguono dalle precedenti nel fatto che sono cicliche. Il sistema di tali reti è dinamico, dato un determinato stimolo la risposta della rete non viene dettata soltanto dai caratteri strutturali della rete stessa, ma varia in funzione del precedente contesto in cui si è manifestato lo stimolo. L’output non è determinato solo dall’input, ma anche da una cronologia di input che fornisce una forma di memoria a breve termine.

*Inizio capitolo DUE*

**Slide 9: Addestramento e paradigmi**

Per costruire una rete neurale eﬃciente un passo fondamentale è individuare un insieme di apprendimento detto training set ovvero una collezione di esempi dal quale la rete attingere per raggiungere il suo scopo ed un algoritmo di apprendimento ovvero un procedimento che permetta di prelevare le informazioni dal training set e di ﬁssare dei parametri che vengano poi modiﬁcati attraverso operazioni iterative interfaccianti con l’ambiente. L’apprendimento avviene sempre grazie ad un certo numero di esempi prelevati dal mondo reale ed opera in due fasi distinte:

• Fase di apprendimento o addestramento nota come learning durante la quale si cerca di far imparare alla rete tutte le informazioni contenute nel training set ottenendo un modello che verrà poi utilizzato nella fase di generalizzazione per analizzare nuovi ingressi

• Fase di generalizzazione nota come recall dove per generalizzazione di una rete si intende la sua capacità di fornire le risposte appropriate a pattern di input che non sono mai stati incontrati e quindi in questi termini si dice che una rete generalizza bene se produce mappature input/output corrette anche se l’input è lievemente diﬀerente dagli esempi usati in fase di training.

Le reti neurali si ispirano al tratto caratteristico del sistema nervoso, ovvero la capacità di acquisire esperienza da esempi del mondo reale: per questo, oltre che di apprendimento, si parla di addestramento delle reti neurali attraverso dei paradigmi di apprendimento.

**Slide 10: Paradigmi di apprendimento (superv e non superv)**

I paradigmi di apprendimento si suddividono in:

• apprendimento supervisionato che presuppone un training set nel quale sono presenti coppie del tipo (xk,ydk) dove la prima variabile indica il k-esimo ingresso e la seconda la k-esima uscita desiderata. Con yk si indica l’uscita reale e la si confronta con l’uscita desiderata: l’obiettivo è modiﬁcare i pesi aﬃnchè si minimizzi la diﬀerenza tra le due uscite. Il training set iniziale viene proposto ripetutamente ﬁnch´e yk ≈ ydk, ovvero l’uscita reale sia il piu` simile possibile a quella desiderata, il tutto modiﬁcando i pesi in base alla legge di apprendimento scelta.

• apprendimento non supervisionato durante il quale la rete modiﬁca i pesi autonomamente quindi si auto-organizza. Viene fornito solo il training set senza precisare le uscite.

**Slide 11: Paradigmi di apprendimento**

I processi comunemente usati sono:

* Delta rule: regola che modiﬁca in maniera proporzionale solo i pesi delle connessioni che hanno contribuito all’errore. Indicando con e il vettore k-esimo degli ingressi ed il vettore dei pesi al k-esimo ingresso e con ed l’uscita ottenuta e l’uscita desiderata, si definisce l’errore il valore . Se consideriamo la variazione del generico vettore dei pesi si ha che: . Quindi l’algoritmo è caratterizzato da questi due passi: se si ottiene che l’uscita ottenuta corrisponde con l’uscita desiderata nessuna modifica dei pesi; nel caso contrario la modifica sui pesi avviene seguendo .
* Discesa del gradiente: Il segnale di errore definito anche nella delta rule dà il via ad un meccanismo di controllo che va ad applicare una sequenza di modiﬁche ai pesi sinaptici del neurone interessato al ﬁne di avvicinare la risposta ottenuta a quella desiderata. Tale processo di ricerca dei pesi migliori si basa sulla scelta di pesi che minimizzano una funzione errore E(), costruita al variare dei pesi stessi. Per la scelta di tali pesi si sfruttano le informazioni fornite dal gradiente locale della funzione errore costruita. Se con indichiamo il gradiente della funzione, allora si ha che e quindi la ricerca risulta essere guidata in modo proporzionale. Calcolando il gradiente della funzione e facendo le opportune sostituzioni si avrà con g funzione di attivazione.
* Hebbian: Indicando il vettore k-esimo degli ingressi ed il vettore dei pesi al k-esimo ingresso l’uscita ottenuta e l’uscita desiderata con la notazione utilizzata per la delta rule, nell’ipotesi che uscita desiderata ed ottenuta non coincidano viene modiﬁcato ogni valore del peso di una connessione seguendo questo schema: se allora , invece se si fa con sempre il learning rate
* Competitivo: infine questo ultimo processo si basa su una vera a propria competizione tra i neuroni di uscita di una rete neurale per attivarsi in seguito ad uno stimolo. In un certo momento t può attivarsi un solo neurone che viene denominato winners-takes-all. Il neurone k che si attiva è quello con input netto più alto per un dato input x, con combinazione lineare di tutti gli input. Il suo segnale di output sarà 1 mentre quello degli altri neuroni rimasti inattivi è 0 ed esso apprende spostando i pesi dagli input inattivi agli input attivi.

**Slide 12: Attività di apprendimento**

La scelta del paradigma di apprendimento, e quindi la metodologia di addestramento di una rete neurale, avviene in base all’attività che la rete deve svolgere ovvero il problema che deve risolvere.

* Memoria associativa: Per memoria associativa intendiamo un criterio di memorizzazione e recupero di informazioni attraverso l’associazione. Completamente ispirata al metodo di memorizzazione del cervello umano, consente il recupero dell’informazione sulle basi di una conoscenza parziale del suo contenuto senza conoscerne la locazione di memoria. La memoria associativa si articola in due fasi, la fase di memorizzazione in cui la rete viene addestrata per fare in modo che dei pattern, vettori di numeri reali, vengano memorizzati ed associati e la fase di richiamo durante la quale si richiama dalla rete un pattern memorizzato a seguito della presentazione di una versione parziale o distorta di un pattern chiave.
* Riconoscimento di pattern: in una sessione di addestramento alla rete vengono presentati ripetutamente un insieme di pattern, lo spazio dei dati, speciﬁcando per ognuno la classe di appartenenza. In seguito alla rete viene presentato un pattern mai mostrato precedentemente, ma appartenente ad una categoria di pattern che la rete ha memorizzato. La rete neurale sarà in grado di riconoscerne la categoria e classiﬁcare il pattern grazie ai dati precedentemente acquisiti. Il processo di riconoscimento inizia con l’estrazione delle features ovvero le caratteristiche che un generico pattern nello spazio dei dati deve rispettare. Poi avviene un’estrazione, ovvero una trasformazione che collega il pattern x con un punto y

nello spazio delle caratteristiche. Inﬁne un’ultima trasformazione detta classiﬁcazione mappa dal punto y ad una regione dello spazio decisionale che è uno spazio multidimensiolae suddiviso in regioni determinate dalla rete nella fase di addestramento, ognuna delle quali è associata ad una classe.

* Infine Approssimazione di funzioni: Si prenda quindi in esame una rappresentazione non lineare input/output descritta da y = f(x) dove x è l’input e y è l’output e si consideri un insieme di esempi (xi,yi) per i = 1,2,...,n. L’incognita è la funzione f ed il problema consiste nel creare una rete capace di generare una rappresentazione F quanto più vicina alla funzione f. In senso euclidiano bisogna soddisfare la condizione ||F(x)−f(x) < ε|| ∀x dove ε è un numero positivo piccolo, che risulterà essere sempre più piccolo con l’accrescere delle dimensioni e dei parametri liberi della rete.

**Slide 13: Cosa sono le reti di Hopfield**

Le reti di Hopﬁeld furono presentate nell’anno 1982 dal ﬁsico John Hopﬁeld in un suo famoso articolo. Hopﬁeld cercò di fare emergere gli schemi di funzionamento che erano dietro ai comportamenti biologici elementari ed il modello matematico realizzato da Hopﬁeld dimostrò che un gran numero di semplici elementi di elaborazione possedevano capacità computazionali dal comportamento collettivo e che tali capacità emergevano spontaneamente come conseguenza di interazioni tra le unità.

Queste reti sono utilizzate principalmente nella costruzione di memorie associative per il riconoscimento di rappresentazioni alterate ed il recupero di informazioni mancanti, imitando quindi la capacità del nostro cervello di riconoscere un’immagine anche quando questa non è esatta, risulta deformata o distorta e di riuscire a leggere parole con lettere mancanti. Tali reti sono impiegate anche:

• nella risoluzione di problemi d’ottimizzazione combinatoriale;

• nei problemi di classiﬁcazione.

**Slide 14: Architettura ed aggiornamento**

Sono reti completamente connesse: Ogni unità è collegata a tutte le altre attraverso connessioni simmetriche, in termini di pesi sinaptici si ha .

Non si hanno auto-connessioni e quindi e

la matrice dei pesi sinaptici è simmetrica con funzione di attivazione a gradino o a soglia.

I valori di input di ogni unità sono binari e non c’è distinzione tra unità di input ed unità di output.

Nella fase di apprendimento la rete memorizza un certo numero di informazioni dette stati stabili, mentre nella fase operativa modiﬁca il proprio stato sulla base dei pesi colleganti i vari neuroni: la rete itera ﬁno a convergere su dei valori di output che rappresentano il pattern più simile all’input memorizzato dalla rete.

La rete di Hopﬁeld è una rete ricorrente e ad ogni rete con una simile architettura è associata una sequenza di stati. Nel nostro caso tale sequenza è rappresentata da un vettore che coincide col vettore delle attività dei neuroni.

Detto lo stato della rete all’istante , lo stato successivo viene calcolato come: con W matrice dei pesi.

Proposito: L’idea di Hopﬁeld fu quella di attribuire alla multi-stabilità della struttura una memoria. Ogni stato stabile può essere visto come un attrattore per gli stati iniziali vicini ad esso, ovvero un insieme verso il quale un sistema dinamico evolve dopo un tempo suﬃcientemente lungo. Inizializzando quindi la rete ad uno stato iniziale, quest’ultimo evolverà verso l’attrattore. Il proposito principale `e fare in modo che certi stati siano appresi come stabili e con modiﬁche opportune sui pesi siano resi attrattori.

**Slide 15: Stabilità e funzione energia**

Ma nelle reti ricorrenti non è detto a priori che la rete riesca a convergere su uno stato stabile. Hopﬁeld ha veriﬁcato che la stabilità del sistema delle sue reti è strettamente legata ad una funzione chiamata funzione energia della rete che indichiamo con E. Aggiungendo l’ipotesi che la matrice dei pesi sia simmetrica con diagonale principale nulla e che si lavori in una condizione di aggiornamento asincrono, Hopﬁeld ha dimostrato che E non può aumentare nel passaggio da uno stato all’altro e poichè non può raggiungere valori inﬁnitamente negativi, lo stato della rete cesserà di cambiare in un tempo ﬁnito ed essa raggiungerà dunque uno stato stabile. Infatti generalmente se consideriamo un sistema dinamico, ad una sua condizione di equilibrio stabile corrisponde un minimo dell’energia posseduta dal sistema stesso come conseguenza del fatto che esso tende spontaneamente a minimizzarla. L’energia del sistema viene espressa come una funzione di Lyapunov veriﬁcante il teorema

Teorema 1 (Lyapunov). Lo stato di equilibrio x è stabile se in un piccolo intorno di x esiste una funzione scalare deﬁnita positiva V (x) tale che la sua derivata rispetto al tempo è minore o uguale a 0 in quella regione.

Quindi in una rete di Hopfield gli stati stabili coincideranno con i minimi della funzione energia che risulterà essere una funzione di Lyapunov e questa risulterà essere convergente proprio in essi. Questi funzioneranno pertanto come bacini d’attrazione per gli stati della rete.

La rete di Hopﬁeld può essere quindi vista come un sistema dinamico non lineare che evolve verso una conﬁgurazione stabile e la stabilità è legata alla funzione energia.

**Slide 16: Caso discreto**

In questo modello ogni neurone può assumere due stati e . Lo stato istantaneo del sistema è speciﬁcato elencando gli n valori di vi e lo stato di ciascun neurone ad un certo istante è determinato da una formula di attivazione binaria

dove e la soglia di attivazione associata al neurone i-esimo**.**

La funzione energia assume questa forma

E risulta essere limitata inferiormente, strettamente monotona decrescente lungo una qualsiasi traiettoria descritta dalla rete di Hopﬁeld nello spazio di stato: