

Chiara Macchiavello

**Dipartimento di Fisica A. Volta
Università di Pavia**

INTRODUZIONE ALLE RETI NEURALI

Seminario tenuto il 17 Dicembre 1992

1 Introduzione

Giorno per giorno, soprattutto in ambiente scientifico, occorre confrontarsi con due distinte classi di problemi. La prima comprende problemi di calcolo numerico, che non siamo in grado di trattare a mano, ma che con i calcolatori elettronici sono di veloce risoluzione, in quanto facilmente descrivibili da algoritmi sequenziali. La seconda invece è costituita da problemi più difficilmente traducibili in termini matematici, quali ad esempio il riconoscimento di immagini o di suoni, che il cervello umano è in grado di risolvere molto velocemente, a differenza dei computer tradizionali. Le reti neurali si propongono come alternativa al calcolo sequenziale in questo secondo caso. L'idea fondamentale consiste infatti nel costruire un meccanismo di elaborazione delle informazioni che si comporti in modo simile al cervello umano. Le caratteristiche del cervello e delle cellule neuronali costituiscono allora il primo criterio su cui basarsi per progettare le reti, come vedremo nel prossimo paragrafo.

I primi tentativi di schematizzare con modelli matematici il cervello umano risalgono a oltre mezzo secolo fa: nel 1943 McCulloch e Pitts propongono un modello di neurone e la connessione di un grande numero di queste unità per la realizzazione di una rete neurale. Successivamente, nel 1957, Rosenblatt formula il Perceptron (descritto nel terzo paragrafo), il primo dispositivo in grado di riconoscere e classificare configurazioni. Nel 1969 Minsky e Papert mettono in luce i limiti del Perceptron e la sua incapacità di risolvere alcuni problemi di base, come ad esempio la funzione logica OR esclusivo. Ciò ha bloccato per circa vent'anni ogni progresso in questo campo. Soltanto verso la metà degli anni '80 si assiste ad una rinascita delle reti neurali, grazie agli studi di Rumelhart, Hinton e Williams, che propongono la formulazione di reti con strati "nascosti" con la tecnica di apprendimento "error back propagation", e ad Hopfield, che mette a punto un tipo di rete neurale ciclica, particolarmente adatta a risolvere problemi di ottimizzazione. Nel quarto e quinto paragrafo vengono descritti tali modelli, che sostanzialmente sono tra i più utilizzati nelle applicazioni pratiche. In conclusione sono riassunti i principali vantaggi e svantaggi che le reti neurali presentano rispetto all'uso dei calcolatori tradizionali.

2 Caratteristiche del cervello umano e caratteristiche di una rete neurale

Il cervello umano è costituito da $10^{10} - 10^{12}$ cellule nervose, chiamate neuroni, connesse tra di loro in vario modo. Il numero di interconnessioni è stimato dell'ordine di 10^4 per ognuna di esse. Il neurone è composto essenzialmente da un corpo cellulare, un assone e numerosi dendriti, come si può osservare in Fig. 1. Il corpo cellu-

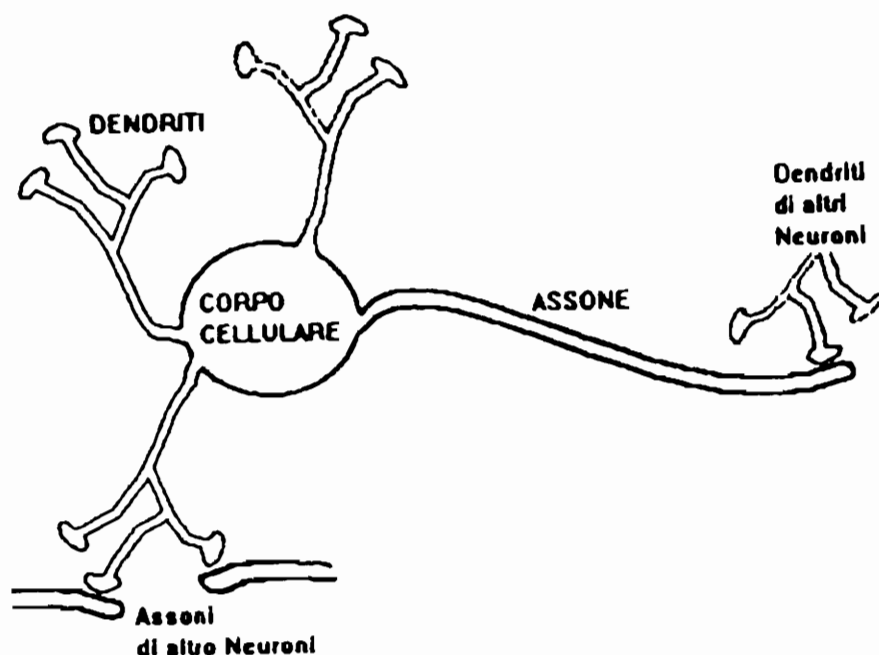


Figura 1. *Il neurone* (da [4]).

lare è rivestito da una membrana in grado di mantenere una certa concentrazione di cariche elettriche, provenienti dai dendriti; questi sono dei prolungamenti ramificati a loro volta in contatto con altri neuroni. I molteplici dendriti di un neurone concorrono ad eccitare il corpo cellulare il quale, *oltre una certa soglia*, invia un impulso elettrico attraverso l'assone a tutti i neuroni cui è connesso. Il punto di contatto tra due neuroni si chiama sinapsi, una leggera intercapedine tra assone e dendrite; il suo spessore può variare nel tempo, provocando quindi un rafforzamento o un indebolimento della connessione tra due neuroni. Il processo di apprendimento del cervello

umano avviene tramite queste modificazioni dell'efficienza con cui le sinapsi trasmettono un segnale da un neurone all'altro.

I neuroni non hanno prestazioni eccezionali: essi sono molto più lenti dei componenti elementari dei calcolatori, essendo i rispettivi tempi di commutazione dell'ordine di 1 msec. contro 10 nanosec. Ciò spiega l'ipotesi di un meccansimo di trasmissione dell'informazione di tipo parallelo, contrariamente alla natura sequenziale degli ordinari calcolatori elettronici. L'elaborazione risiede cioè in intere regioni del cervello piuttosto che in singole unità e la realizzazione di comportamenti complessi emerge dalla cooperazione di moltissime unità elementari di per sè semplici, ma riccamente interconnesse tra di loro e operanti in parallelo. Inoltre, mentre nei computer elettronici l'elaborazione viene effettuata nell'unità centrale di calcolo (CPU) e la memoria è concentrata in unità separate (RAM), nel cervello non vi è netta distinzione fra memoria e unità di elaborazione.

Queste caratteristiche del cervello umano ci suggeriscono i primi criteri basilari per progettare una rete neurale:

- deve essere costituita da un numero molto elevato di componenti elementari, in grado di svolgere solo operazioni semplici (esecuzione di medie pesate e decisioni di soglia);
- ogni componente deve essere interconnesso con un numero elevato di altri componenti;
- le connessioni tra i vari componenti non devono essere rigide, ma modificabili mediante opportuni processi di apprendimento.

Il processo di apprendimento assume un'importanza fondamentale per il buon funzionamento di una rete neurale. Non è infatti possibile, in generale, calcolare a priori i pesi delle connessioni tra i vari componenti (l'equivalente delle sinapsi) in funzione del compito che la rete deve svolgere. Tali pesi dovranno essere appresi; in particolare, il processo di apprendimento rinforza selettivamente certe connessioni o crea nuove connessioni, ne indebolisce o elimina altre, a seconda delle prestazioni conseguite dalla rete.

Il primo modello matematico di neurone (schematizzato in Fig. 2) è stato proposto nel 1943 da McCulloch e Pitts ed è sostanzialmente tuttora in uso. Lo stato di eccitazione del neurone i -esimo,

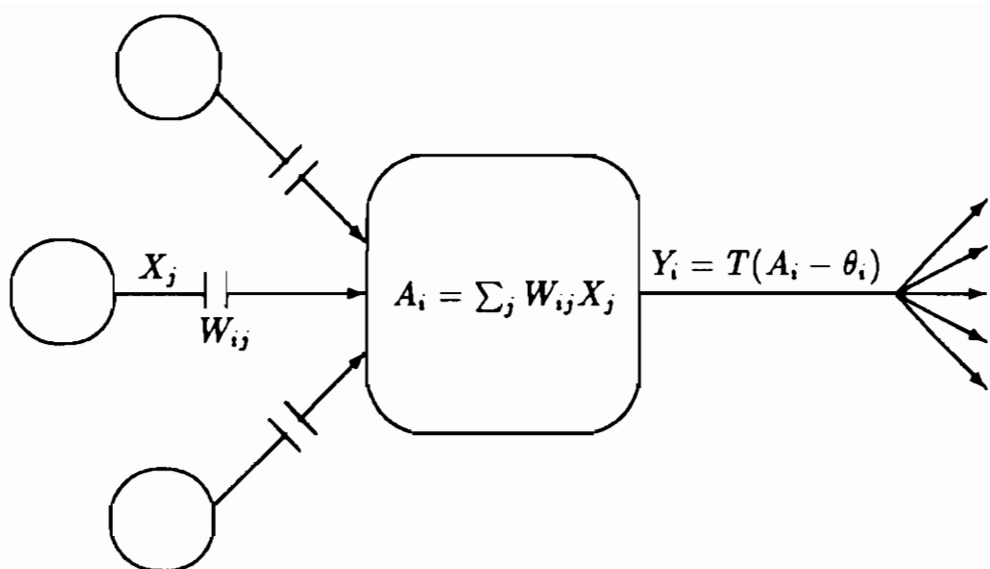


Figura 2.

che riceve segnali provenienti da n neuroni (o unità), è determinato da un valore di attivazione A_i della forma

$$A_i = \sum_{j=1}^n W_{ij} X_j , \quad (1)$$

dove W_{ij} rappresentano i pesi delle connessioni tra i neuroni e X_j il valore in uscita della j -esima unità. Il meccanismo di scarica dell'attivazione lungo l'assone è rappresentato da una funzione di trasferimento T . Nel modello originale T è la funzione a gradino di Heaviside e le uscite dei neuroni assumono i valori binari 0,1. Il valore di uscita dell' i -esima unità è dato dall'espressione

$$Y_i = T(A_i - \theta_i) , \quad (2)$$

dove θ_i è il valore di soglia, al di sotto del quale il neurone non viene attivato.

Collegando tra loro in vario modo le unità elementari si ottengono diverse configurazioni di rete. Una rete neurale è dunque definita dal numero di unità di cui è composta, dalla particolare

topologia che determina i collegamenti tra le varie unità e dal tipo di funzione di trasferimento utilizzata.

Il funzionamento di una rete è in generale caratterizzato da due fasi ben distinte nel tempo. La prima è quella dell'apprendimento, cui si è già accennato in precedenza; in questa fase la rete viene addestrata fornendo esempi noti, si calcolano cioè i pesi tra le varie unità a partire da valori di input e di output dati. La seconda fase è quella dell'esecuzione vera e propria, in cui la rete elabora dati mai visti prima e fornisce i valori di output a partire da quelli di input e dei pesi (determinati durante l'apprendimento). Esiste in realtà anche un altro tipo di processo di apprendimento, detto "non supervisionato", in cui gli esempi da apprendere non sono associati ad un valore di output dato. Lo scopo delle reti basate su tale meccanismo è infatti quello di classificare gli input ricevuti individuandone possibili categorie non note a priori. In questo lavoro non verrà trattato tale caso.

3 Il Perceptron

Il Perceptron, formulato da Rosenblatt nel 1957 per il riconoscimento e la classificazione di immagini, costituisce il primo modello di rete neurale. Può essere essenzialmente schematizzato come un'unità che elabora n valori di input X_i in modo analogo al neurone di Pitts e McCulloch, utilizzando come funzione di trasferimento la funzione di Heaviside, e fornisce un unico valore di output Y . La sostanziale novità rispetto al semplice modello di Pitts e McCulloch consiste nella possibilità di modificare i pesi dei valori di input e il valore di soglia in base a un ben preciso sistema di addestramento. Per poter riaggiustare la soglia in modo analogo ai pesi è utile introdurre un nuovo dato di input X_0 , che assume sempre valore unitario, e un nuovo peso $W_0 = -\theta$, pari al valore della soglia. In questo modo la funzione di attivazione assume la forma

$$A = \sum_{i=0}^n W_i X_i, \quad (3)$$

viene cioè diminuita del valore di soglia.

Il processo di apprendimento consiste nel partire da valori casuali dei pesi e della soglia e fornire determinati esempi noti, che devono essere correttamente memorizzati variando in modo opportuno i pesi e la soglia. Più precisamente, per ogni esempio da apprendere si calcola il valore di output Y ; se $Y = 0$ e dovrebbe essere 1 si pone

$$W_i(t+1) = W_i(t) + X_i, \quad (4)$$

cioè si incrementano o si lasciano inalterati i vari pesi, a seconda del rispettivo valore X_i , in modo da aumentare l'attivazione. Se invece si trova $Y = 1$ mentre il risultato voluto è 0, i pesi vengono diminuiti

$$W_i(t+1) = W_i(t) - X_i. \quad (5)$$

Questo processo viene iterato finché si ottiene il risultato corretto e si esauriscono gli esempi da apprendere. A questo punto i pesi vengono fissati e si può procedere alla fase di esecuzione, in cui nuovi dati vengono riconosciuti o classificati sulla base di quelli memorizzati. Si può dimostrare che questo processo di apprendimento, se la soluzione esiste, converge sempre.

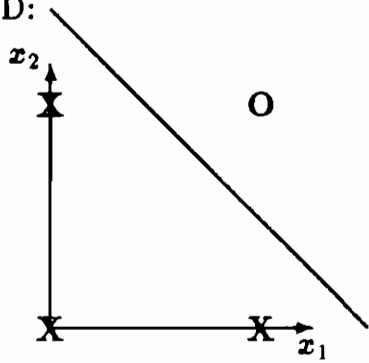
A causa della natura della funzione di trasferimento e dell'attivazione, il Perceptron presenta una grossa limitazione. Il valore di output fornito infatti è 0 oppure 1 a seconda che la combinazione lineare $\sum_i W_i X_i$ sia negativa o positiva. La scelta dei pesi individua cioè un iperpiano a $n - 1$ dimensioni nello spazio n -dimensionale dei valori di input; *è possibile pertanto risolvere solo problemi linearmente separabili*.

Per chiarire meglio questo limite consideriamo il caso semplice di input bidimensionali: come illustrato in Fig. 3, è facile risolvere il problema delle forme logiche AND e OR in quanto le due classi di valori di output sono separabili da una retta, mentre il caso della forma logica XOR (or esclusivo) rappresenta un problema non linearmente separabile e non può quindi essere risolto correttamente dal Perceptron.

I limiti del Perceptron sono d'altra parte comprensibili se si pensa all'estrema semplicità del modello, che risiede sia nella topologia di rete (un solo livello decisionale), sia nella singola unità (solo valori binari con funzione di trasferimento a gradino), sia nella

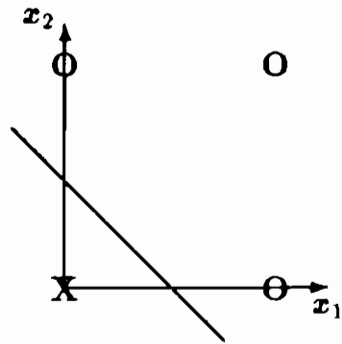
Funzione AND:

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1



Funzione OR:

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1



Funzione XOR:

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

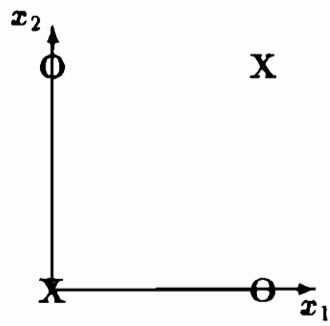


Figura 3.

legge di apprendimento estremamente semplificata. Come vedremo nel prossimo paragrafo queste limitazioni possono essere superate costruendo reti più sofisticate.

4 Reti neurali a più strati

Le reti neurali a più strati sono caratterizzate da un insieme di unità di input, un insieme di unità di output e da uno o più strati di unità “nascoste” che collegano il livello di input a quello di output, come si può osservare in Fig. 4. Le connessioni tra i livelli

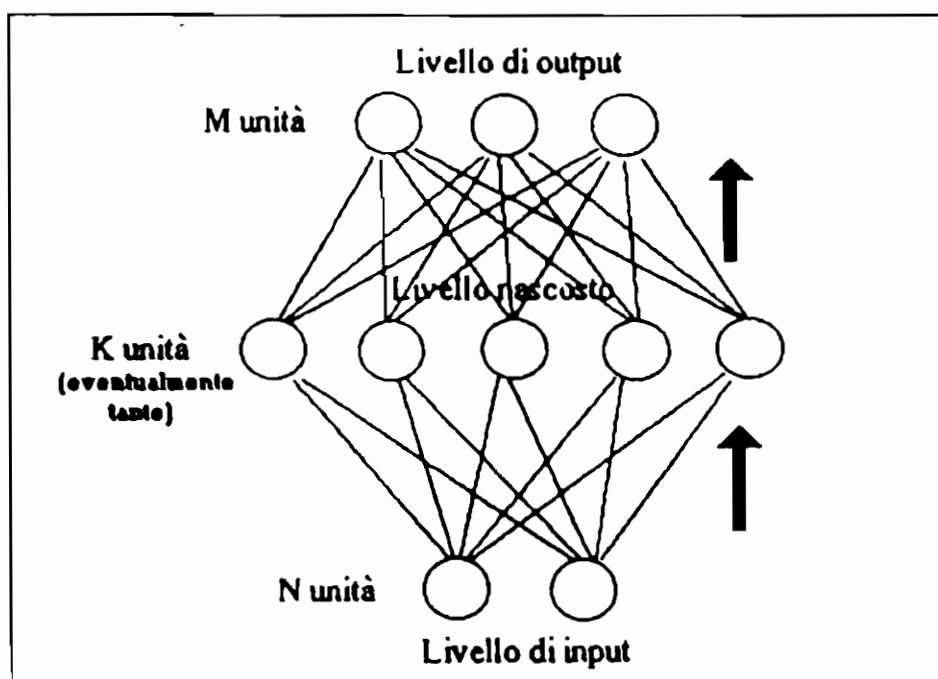


Figura 4. Rete a più strati (da [4])

sono totali e in una sola direzione: ogni unità riceve un segnale da tutte le unità dello strato precedente e trasmette il proprio valore di uscita, opportunamente pesato, a tutte le unità dello strato successivo. La struttura delle singole unità è un po' più complessa rispetto al modello originale di neurone: l'uscita è rappresentata da un valore reale compreso nell'intervallo $[0,1]$ e la funzione di trasferimento può essere una qualsiasi funzione monotona differenziabile.

Queste caratteristiche sono richieste dalla particolare procedura di apprendimento di queste reti, come vedremo in seguito.

Questo semplice tipo di rete è interessante perchè le unità nascoste possono costruire liberamente le proprie rappresentazioni dei segnali in ingresso, definendo opportunamente i vari pesi. In questo modo si possono risolvere anche problemi più complessi rispetto al Perceptron. Si può infatti dimostrare che una qualsiasi funzione può essere calcolata da una rete neurale con tre strati decisionali, cioè con uno strato di uscita e due strati nascosti (Teorema di Kolmogorov). In particolare, reti con uno strato decisionale, come il Perceptron, possono risolvere problemi di separazione lineare; reti con due strati decisionali sono in grado di separare lo spazio dei valori in ingresso in semi-spazi convessi, mentre sono sufficienti tre strati decisionali per individuare semi-spazi di forma qualsiasi (v. Fig. 5).

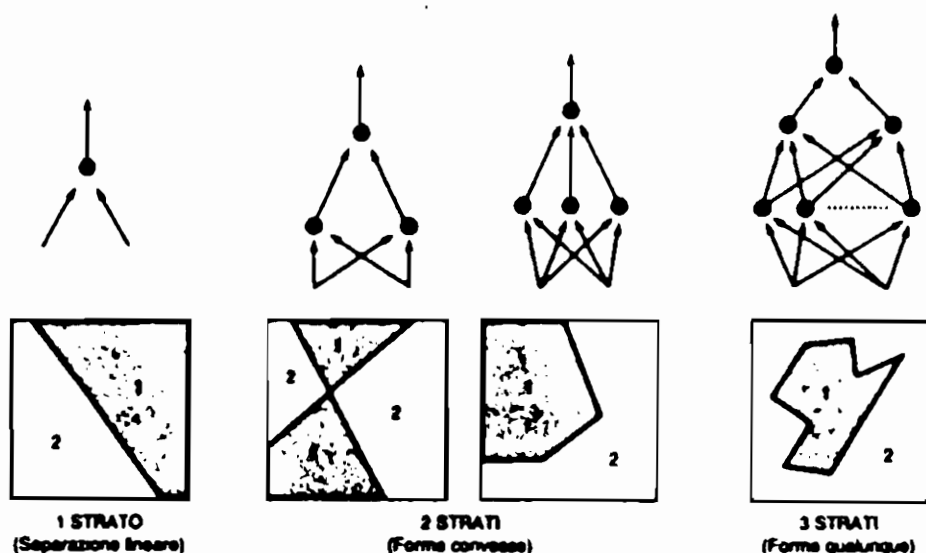


Figura 5 (da [3])

Il problema della funzione logica XOR si risolve facilmente individuando nello spazio bidimensionale degli input una striscia che ricopra i due valori nulli oppure i due valori unitari e pertanto occorre aggiungere un solo strato di unità nascoste al Perceptron.

Nei casi pratici il numero di unità di input e quello di output sono determinati dal problema che si vuole risolvere, mentre purtroppo non esistono criteri rigorosi per stabilire il numero ottimale degli strati intermedi e quello delle loro unità. In generale, comunque, non si adottano più di uno o due strati intermedi e il numero dei loro neuroni è spesso inferiore sia a quello degli input che a quello degli output.

Occupiamoci ora del processo di apprendimento, che, come è già stato sottolineato, riveste un ruolo di fondamentale importanza per il buon funzionamento della rete. In genere l'aggiornamento dei pesi delle connessioni di ogni unità viene effettuato in modo da minimizzare l'errore commesso da questa (cioè la differenza tra il valore di uscita calcolato e quello voluto). In questo tipo di rete però l'uscita voluta non è nota per le unità nascoste e non si conosce quindi a priori un criterio per modificarne i relativi pesi. Questo problema è stato risolto con la tecnica dell'*error back propagation* da Rumelhart, Hinton e Williams nel 1986. L'errore totale commesso dalla rete sia

$$E = \sum_k E_k = \frac{1}{2} \sum_k \sum_j (D_{kj} - Y_{kj})^2, \quad (6)$$

dove k rappresenta l'insieme degli esempi noti che la rete deve apprendere, mentre D_{kj} e Y_{kj} sono rispettivamente il valore di uscita voluto e quello calcolato dall'unità di output j -esima per l'esempio k -esimo. Per minimizzare questo errore viene utilizzata la tecnica di discesa del gradiente: si assegnano i pesi iniziali in modo casuale e li si varia nel tempo secondo la legge

$$\Delta W_{ij} = -\eta \frac{\partial E_k}{\partial W_{ij}}, \quad (7)$$

dove η ($0 < \eta < 1$) è detto tasso di apprendimento. La formula (7) garantisce che tale variazione dei pesi porta a un minimo di E .

Consideriamo innanzitutto i pesi che connettono l'ultimo strato nascosto della rete con lo strato di uscita. In tal caso la variazione ΔW_{ji} relativa alla j -esima unità di uscita e l' i -esima unità dello strato precedente si può scrivere

$$\Delta W_{ji} = -\eta \frac{\partial E_k}{\partial Y_j} \frac{\partial Y_j}{\partial A_j} \frac{\partial A_j}{\partial W_{ji}}. \quad (8)$$

Tenendo conto dell'espressione (1) per l'attivazione A_j e della (6) si ottiene

$$\Delta W_{ji} = -\eta d_j X_i, \quad (9)$$

dove

$$d_j = T'(A_j)(Y_j - D_j) \quad (10)$$

e X_i indica il valore di uscita della i -esima unità dello strato nascosto precedente all'output. Si noti che l'espressione (10) contiene la derivata della funzione di trasferimento: da qui la richiesta di derivabilità introdotta all'inizio del paragrafo.

Per i pesi che connettono l'ultimo strato nascosto con quello precedente si ottiene invece

$$\Delta W_{ir} = -\eta \frac{\partial E_k}{\partial X_i} \frac{\partial X_i}{\partial A_i} \frac{\partial A_i}{\partial W_{ir}} = -\eta \frac{\partial E_k}{\partial X_i} T'(A_i) Z_r. \quad (11)$$

In questo caso la derivata dell'errore rispetto all'uscita dell'unità nascosta non è nota; si può però risalire al suo valore sfruttando i valori noti per lo strato di output: si sviluppa la derivata dell'errore nel modo seguente

$$\frac{\partial E_k}{\partial X_i} = \sum_j \frac{\partial E_k}{\partial Y_j} \frac{\partial Y_j}{\partial A_j} \frac{\partial A_j}{\partial X_i} = \sum_j \frac{\partial E_k}{\partial Y_j} T'(A_j) W_{ji} = \sum_j d_j W_{ji} \quad (12)$$

La (11) diventa allora

$$\Delta W_{ir} = -\eta T'(A_i) Z_r \left[\sum_j d_j W_{ji} \right]. \quad (13)$$

Combinando questi risultati, la formula generale per l'aggiornamento dei pesi W_{ji} tra la j -esima unità di uno strato (s) e l' i -esima unità dello strato precedente ($s-1$) assume la forma

$$\Delta W_{ji} = -\eta d_j^{(s)} O_i^{(s-1)}, \quad (14)$$

dove O_i è il valore di uscita dell' i -esima unità e

$$d_j^{(s)} = (O_j^{(s)} - D_j^{(s)}) T'(A_j^{(s)}) \quad (15)$$

se s è lo strato di output, mentre

$$d_j^{(s)} = T'(A_j^{(s)}) \sum_r d_r^{(s+1)} W_{rj} \quad (16)$$

per ogni strato nascosto.

Nella maggior parte dei casi pratici si utilizza come funzione di trasferimento

$$T(A) = \frac{1}{1 + e^{-A}}, \quad (17)$$

che permette una notevole semplificazione dei calcoli, infatti

$$T'(A) = T(A)(1 - T(A)) \quad (18)$$

e

$$d_j^{(s)} = (O_j^{(s)} - D_j^{(s)}) O_j^{(s)} (1 - O_j^{(s)}) \quad (19)$$

$$d_j^{(s)} = O_j^{(s)} (1 - O_j^{(s)}) \sum_r d_r^{(s+1)} W_{rj} \quad (20)$$

rispettivamente nei due casi esposti in precedenza.

L'apprendimento con la tecnica *error back propagation* è di tipo ciclico: per ogni esempio da apprendere si calcolano i valori di attivazione e di uscita dal primo strato nascosto fino ad arrivare allo strato di output. Si calcolano quindi i valori di d_j e le variazioni dei pesi iniziando dallo strato di output e per ogni strato nascosto, partendo dall'ultimo e risalendo all'indietro fino al primo (da cui il nome "back propagation"). Questo processo viene iterato finché E diventa minore di un errore prefissato, idealmente nullo.

Con riferimento allo strato di output, si noti che, se il valore di output calcolato Y_j è 0 oppure 1, si ha sempre $d_j = \Delta W_{ji} = 0$, anche se Y_j fosse diverso da D_j e quindi anche se fosse necessario aggiornare i pesi. Questo è chiaramente un punto debole della procedura di apprendimento che stiamo trattando. Un altro punto debole è la lentezza di questa procedura, che richiede spesso un gran numero di iterazioni prima di conseguire un errore globale sufficientemente piccolo. Per velocizzare la convergenza del processo è utile

in molti casi pratici modificare la legge di aggiornamento dei pesi con l'introduzione di un termine "di momento"

$$W_{ji}(t+1) = W_{ji}(t) + \Delta W_{ji} + \beta (W_{ji}(t) - W_{ji}(t-1)) , \quad (21)$$

dove ΔW_{ji} è la variazione dei pesi calcolata con le formule precedenti, mentre β è una costante positiva < 1 (detta "momento"). L'aggiunta di questo termine, il cui effetto è controllato dal valore assegnato a β , rappresenta una specie di ricordo dell'aggiornamento precedente.

Un altro potenziale inconveniente delle reti a strati nascosti è la presenza di minimi relativi dell'errore, anche se nella pratica l'intrappolamento in minimi relativi non buoni si verifica raramente ed è considerato meno critico della lentezza di convergenza. Questo limite si può comunque migliorare con le seguenti modifiche, che però in genere rallentano ulteriormente il processo (si deciderà di volta in volta a quali tra queste ricorrere a seconda della particolare applicazione cui la rete è destinata):

- riduzione graduale del tasso di apprendimento, in modo che la rete possa inizialmente uscire da eventuali minimi locali ma poi, diminuendo η , rimanga in minimi più profondi senza compiere ampie oscillazioni;
- aumento del numero di unità nascoste, in modo che la rete si costruisca al suo interno una migliore rappresentazione dei dati di input;
- aggiunta di rumore casuale;
- introduzione di effetti del secondo ordine (legati alla derivata seconda delle funzione di trasferimento) nella tecnica di discesa del gradiente.

5 Reti di Hopfield

Le reti di Hopfield hanno una topologia completamente diversa dalle reti a strati. Come schematizzato in Fig. 6, si tratta di reti totalmente connesse (ogni unità è collegata a tutte le altre), le

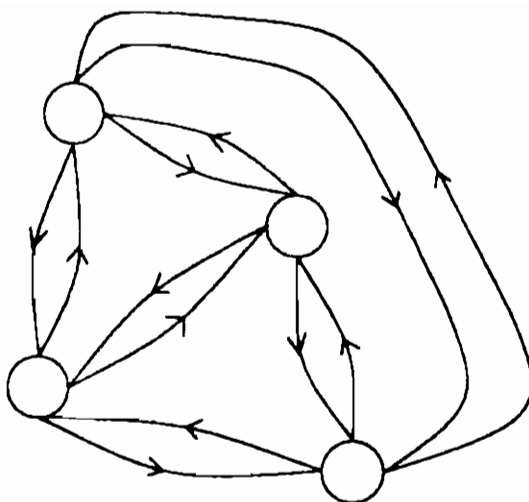


Figura 6. Rete di Hopfield (da [1])

connessioni sono simmetriche ($W_{ij} = W_{ji}$) e non si hanno autoconnessioni ($W_{ii} = 0$). I valori di ogni unità sono bipolari $(-1, +1)$. In questo tipo di reti non c'è distinzione spaziale tra unità di input e unità di output: la rete riceve i valori di input in numero pari alle sue unità e, dopo averli elaborati, li presenta all'uscita attraverso le unità stesse. La regola di apprendimento assume una forma molto semplice: i pesi vengono fissati in un solo stadio, a differenza delle reti a più strati, secondo la seguente formula

$$W_{ij} = \sum_k X_{ki} X_{kj} , \quad (22)$$

dove la sommatoria si riferisce agli esempi che devono essere memorizzati e X_{ik} indica il valore assunto dalla i -esima unità all'esempio k -esimo.

A questo punto la rete può passare alla fase di esecuzione: i valori X_1, \dots, X_n presentati in input vengono modificati come segue

$$X_i(t+1) = +1 \quad (23)$$

se $A_i = \sum_j W_{ij} X_j(t) - \theta_i \geq 0$,

$$X_i = -1 \quad (24)$$

se $A_i = \sum_j W_{ij} X_j(t) - \theta_i < 0$. Tale legge è applicata in modo asincrono, cioè serialmente a un'unità per volta, secondo un ordine prefissato o con scelta casuale. Questo processo viene poi iterato finché le rete si rilassa in uno stato di equilibrio, che corrisponde quindi ai valori di output.

L'evoluzione della rete verso una configurazione stabile è garantita dalle (23) e (24). Hopfield infatti ha dimostrato che alla rete è associata una funzione "energia"

$$E = -\frac{1}{2} \sum_{ij} W_{ij} X_i X_j + \sum_j \theta_j X_j \quad (25)$$

e che le configurazioni relative agli esempi appresi corrispondono ai minimi di E . Durante la fase di esecuzione questa energia diminuisce, infatti

$$\Delta E = \frac{\partial E}{\partial X_i} \Delta X_i = -A_i \Delta X_i \leq 0, \quad (26)$$

dato che, per le (23) e (24), A_i e ΔX_i hanno segno opposto. La configurazione della rete allora si modifica finché si raggiunge lo stato di energia minima. Le reti di Hopfield sono perciò tipicamente utilizzate in problemi di ottimizzazione e di riconoscimento di configurazioni. Un esempio è mostrato in Fig. 7, dove una configurazione "sporca" presentata in input viene via via modificata fino a riprodurre il carattere voluto, che era stato memorizzato dalla rete nella fase di apprendimento. In problemi di questo tipo è possibile memorizzare al massimo $k = 0.14n$ configurazioni diverse (n è il numero totale di unità). Come ordine di grandezza, per memorizzare le lettere dell'alfabeto ($k=21$) è necessario utilizzare una rete con $n \simeq 200$ unità e $\frac{n(n-1)}{2} \simeq 20000$ connessioni diverse.

Uno dei maggiori inconvenienti presentati dalle reti di Hopfield, analogamente al caso delle reti a più strati, è quello della presenza di minimi relativi nell'energia (25). Come soluzione in genere si modifica la dinamica della rete adottando una legge di attivazione di

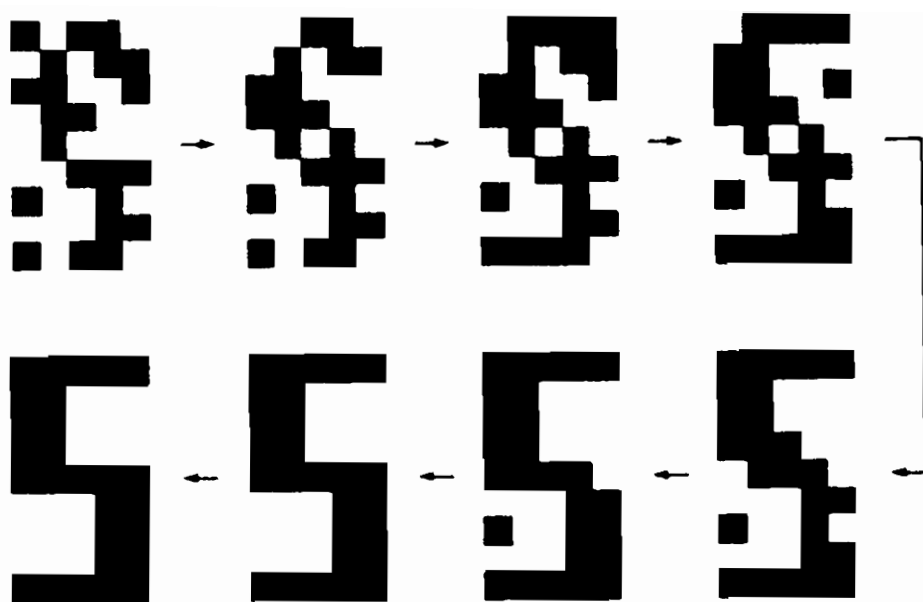


Figura 7. (da [1])

tipo probabilistico, nel senso che ogni unità ha una certa probabilità di non mutare il suo stato anche quando dovrebbe. La probabilità di attivazione della j -esima unità assume la forma

$$P_j(1) = \frac{1}{1 + e^{\frac{-\Delta E_j}{T}}} , \quad (27)$$

dove T rappresenta un parametro di controllo, detto anche temperatura in analogia alla distribuzione termica di Boltzmann. A T viene assegnato inizialmente un valore alto, in modo che la rete possa risalire a valori più alti di energia, superando così eventuali barriere locali; questo valore viene poi diminuito lentamente nel tempo in modo da congelare la rete nella configurazione di energia minima. Lo svantaggio di questa procedura è la lentezza del processo di rilassamento della rete.

6 Conclusioni

Per concludere questa breve introduzione alle reti neurali riassumiamo i vantaggi e gli svantaggi che si presentano nelle applicazioni pratiche.

L'attrattiva fondamentale deriva certamente dalla loro capacità di "comprendere" piuttosto che di memorizzare; questa si manifesta nella possibilità di fare analogie e somiglianze, di elaborare dati non memorizzati e generalizzare gli esempi appresi. Altri due vantaggi di notevole interesse sono la resistenza al rumore e la resistenza al degrado. Resistenza al rumore significa che la rete è in grado di trattare dati incerti, incompleti o "rumorosi" (la capacità di riconoscere input incompleti o rumorosi è detta anche memoria associativa); resistenza al degrado indica la capacità di operare anche in presenza di guasti o parti difettose (quest'ultima caratteristica ha un senso solo nel caso di reti neurali hardware realizzate direttamente con componenti elettronici). Entrambe queste caratteristiche derivano dal fatto che l'informazione viene elaborata in maniera distribuita in una moltitudine di unità elementari ed emerge quindi come fenomeno collettivo; di conseguenza un difetto parziale nell'informazione viene sopprasseduto. Si noti la differenza rispetto ai computer tradizionali, estremamente sensibili al minimo errore. Un altro vantaggio consiste nella velocità computazionale della fase di esecuzione (poichè i calcoli da eseguire sono semplici somme pesate e decisioni di soglia).

Il processo di apprendimento invece richiede in genere tempi molto lunghi e spesso è necessaria una notevole e significativa casistica di esempi, non sempre disponibile o facile da raccogliere. Altre caratteristiche negative sono la scarsa precisione dei risultati ottenuti e la mancanza di una teoria rigorosa nella scelta della rete più adatta al problema in esame, che indichi quanti strati di unità adottare, quante unità per ciascuno strato, il tipo di funzione di trasferimento, ecc.

Le reti neurali costituiscono allora una valida alternativa ai calcolatori tradizionali quando i tempi di esecuzione sono lunghi, non è richiesta una grande precisione ed è preponderante l'esigenza di adattarsi flessibilmente a situazioni mutevoli o difficilmente rappresentabili in forma matematica.

Riferimenti

- [1] R. Beale e T. Jackson, *Neural computing: an introduction*, IOP Publishing, 1990
- [2] G. E. Hinton, *Le Scienze* **291** (1992), 117
- [3] S. Cammarata, *Reti neuronali*, Ed. Bompiani, 1990
- [4] A. Mazzetti, *Reti neurali artificiali*, Ed. Apogeo, 1991
- [5] C. Peterson e T. Rognvaldsson, 1991 Cern school of computing (Proceedings), CERN 92-02, Ginevra 1992

Volume I - n. 1 4/3/1986	CINQUANTA ANNI DI INTERAZIONI DEBOLI: DALLA TEORIA DI FERMÍ ALLA SCOPERTA DEI BOSONI PESANTI - Marcello Conversi	pag. 1
Volume I - n. 2 1/7/1986	EFFECTS OF DIOXINS ON NATURE AND SOCIETY - Opening talk, Sergio P. Ratti	pag. 3
	DIOXIN IN MISSOURI - Armon F. Yanders	pag. 11
	DEMONSTRATION OF INNOVATIVE REMEDIAL ACTION TECHNOLOGIES AT UNITED STATES MILITARY DIOXIN CONTAMINATED SITES - Terry L. Stoddard	pag. 23
	TIMES BEACH DIOXIN RESEARCH FACTLY - Robert J. Schreiber	pag. 41
	E.P.A. RISK ASSESSMENT OF CHLORINATED DIBENZO-P-DIOXIN AND DIBENZOFURANS (CCDs/CDFs) - Donald G. Barnes, Patricia Roberts	pag. 51
	RECENT INTERNATIONAL COOPERATION IN EXCHANGE OF INFORMATION ON DIOXIN - Donald G. Barnes	pag. 63
Volume II - n. 1 15/9/1987	CHLORACNE AND THE AGENT ORAGE PROBLEM IN THE U.S.A. - B. Fischmann	pag. 69
	CONVEGNO SU "LA CONOSCENZA ATTUALE DELLA INTERAZIONE GRAVITAZIONALE" - MOTIVAZIONI DEL CONVEGNO - Sergio P. Ratti	pag. 3
	LA CONOSCENZA ATTUALE DELLA INTERAZIONE GRAVITAZIONALE: UN PROBLEMA APERTO - Sergio P. Ratti, Roberto Silvotti	pag. 5
	SVILUPPI RECENTI SULLA CONOSCENZA DELLA COSTANTE DI GRAVITAZIONE UNIVERSALE - Anna Grassi, Giuliano Strini	pag. 19
	LIMITI SPERIMENTALI SULLA MISURA DELL'ACCELERAZIONE DI GRAVITA' - Roberto Cassinia	pag. 31
	CONSEGUENZE SPERIMENTALI DELLA IPOTESI DI ESISTENZA DI UNA QUINTA INTERAZIONE - Fabrizio Massa	pag. 43
	VERIFICA DEL PRINCIPIO DI EQUITVALENZA E FORZE TRA PARTICELLE ELEMENTARI - Bruno Bertoni	pag. 81
Volume II - n. 2 10/12/1987	TRANSIZIONE LIQUIDO SOLIDO - Mario Tosi	pag. 3
	EQUAZIONI DI MAXWELL NEL VUOTO ED ELETTRODINAMICA QUANTISTICA - Emilio Zavattini	pag. 27
Volume III - n. 1 6/6/1988	METODI DI DILATAZIONE ANALITICA E RISONANZE IN SISTEMI QUANTISTICI NON RELATIVISTICI - Fausto Borgonovi	pag. 1
	CAMPO ELETTRICO ED EMISSIONI DA CARICHE IN UN MEZZO - Michele Spada	pag. 13
	SPETTROSCOPIA VIBRAZIONALE DI SUPERRETICOLI SEMICONDUTTORI - Luciano Colombo	pag. 29
	SOLITONI IN FISICA NUCLEARE - Marco Radici	pag. 51
Volume III - n. 2 4/7/1988	ASPETTI NON LOCALI DEL COMPORTAMENTO QUANTISTICO - Oreste Nicosini	pag. 83
	CARATTERIZZAZIONE OTTICA IN SITU DI FILMS SOTTILI - Alessandra Piaggi	pag. 1
	TRANSIZIONI DI WETTING - Tommaso Bellini	pag. 23
	FORZE A TRE CORPI NEI GAS RARI - Silvia Celi	pag. 49
Volume III - n. 3 15/12/1988	FLAVOUR PHYSICS - Luciano Maiani	pag. 1
	THE STANDARD ELECTROWEAK MODEL: PRESENT EXPERIMENTAL STATUS - Pierre Darriulat	pag. 27
	WHY BE EVEN-HANDED? - Martin M. Block	pag. 47
Volume IV - n. 1 6/4/1989	LA FISICA DEI COLLIDER - Paolo Bagnaia, Fernanda Pastore	pag. 1
Volume IV - n. 2 15/6/1989	SOLAR WIND AND PHYSICS OF THE HELIOSPHERE - Bruno Coppi	pag. 2
Volume IV - n. 3 15/9/1989	THE IGNITOR PROJECT - Bruno Coppi, Francesco Pegoraro	pag. 32
	SPETTROSCOPIA ELLISSOMETRICA NEI SOLIDI - Alessandra Piaggi	pag. 3
	UNA INTRODUZIONE AL SUPERCONDUCTING SUPERCOLLIDER - R. Diaferia	pag. 23
	FENOMENI DI TRASPORTO IN SISTEMI HAMILTONIANI - Fausto Borgonovi	pag. 49
Volume V - n. 1 15/3/1990	MULTI-BODIED PHASE SPACE - A NEW APPROACH - Martin M. Block	pag. 1
	SCATTERING BRILLOUIN RISONANTE - Cristina Bosio	pag. 20
	METODO DI RINORMALIZZAZIONE PER LO STUDIO DELLA STRUTTURA ELETTRONICA DI SUPERRETICOLI - Saverio Moroni	pag. 40
	MECCANISMI DI CONDUCTIBILITA' IONICA MEDIATI DA DIFETTI ESTRINSECI - IL CASO DEL QUARZO - Alberto Paleari	pag. 57
Volume V - n. 2 15/6/1990	STATISTICHE QUANTISTICHE ED INDISTINGUITABILITA' - Gianluca Introzzi	pag. 69
	FISICA DELLE ALTE ENERGIE ALLE KOAN FACTORIES - Renato Diaferia	pag. 1
	NEUTRONI FREDDI E NEUTRONI ULTRAFREDDI - Gianluca Raselli	pag. 15
	TRANSIZIONI ORDINE-DISORDINE NELLE DISPERSIONI COLLOIDALI - Paolo Di Trapani	pag. 51
Volume VI - n. 1 p. I - 11/10/1991	LE UNITA' DI MISURA DELLA RADIOPROTEZIONE - Sergio P. Ratti	pag. 1
	L'INCIDENTE NUCLEARE DI CHERNOBYL - Giuseppe Belli	pag. 7
	NORMATIVA E PRINCIPI DI RADIOPROTEZIONE - Argeo Benco	pag. 17
	APPENDICE 1 - Pubblicazioni I.C.R.P. - Argeo Benco	pag. 56
	APPENDICE 2a) - Il regime giuridico dell'impiego pacifico dell'energia nucleare	pag. 59
	APPENDICE 2b) - Elenco di provvedimenti di interesse per le attività di impiego pacifico dell'energia nucleare e delle radiazioni ionizzanti	pag. 64
	APPENDICE 2c) - Raccolta di Circolari Ministeriali relative all'utilizzazione delle sostanze radioattive e delle macchine radiogene	pag. 78

	APPENDICE 2d) - Raccolta di Circolari Ministeriali relative al trasporto delle sostanze radioattive	pag. 85
	RADIOATTIVITA' AMBIENTALE E RADIOECOLOGICA - Arrigo Cigna	pag. 87
	EFFETTI BIOLOGICI DELLE RADIAZIONI IONIZZANTI - Marco Caproni	pag. 107
Volume VI - n. 1 p. II - 11/10/1991	MODELLO PREVISIONALE DELLA CONCENTRAZIONE DI ^{90}Sr , ^{134}Cs E ^{137}Cs NELLA CATENA ALIMENTARE - Arrigo Cigna	pag. 117
	L'AMBIENTE E LA RADIOPROTEZIONE IN RELAZIONE AD INCIDENTI NUCLEARI - Arrigo Cigna	pag. 131
	INCIDENTE NUCLEARE "CHERNOBYL" E SUE RIPERCUSSIONI SULLA CATENA ALIMENTARE - R. Cazzaniga, G. Dominici, A. Malvicini, E. Sangalli	pag. 157
	PRIMA VALUTAZIONE DELL'IMPATTO RADIOLOGICO AMBIENTALE NELLA ZONA DI ISPRÀ IN RELAZIONE ALL'INCIDENTE NUCLEARE DI CHERNOBYL - Argeo Benco	pag. 177
	APPROCCIO FRATTALE ALLA DESCRIZIONE DELLA RADIOATTIVITA' IN ARIA IN ITALIA DOPO CHERNOBYL - Gianfausto Salvadori	pag. 201
Volume VII - n. 1 15/3/1992	ELECTRON ENERGY LOSS SPECTROSCOPY - Marco Amioti	pag. 1
	LIVELLI ELETTRONICI PROFONDI IN SEMICONDUTTORI E LORO CARATTERIZZAZIONE - Adele Sassella	pag. 35
	LA RICERCA DEL BOSONE DI HIGGS AI FUTURI ACCELERATORI - G. Montagna	pag. 57
Volume VII - n. 2 15/10/1992	SIMMETRIA CHIRALE E TEOREMA DI GOLDBERGER-TREIMAN - Carlo Gobbi	pag. 81
	CRESCITA, CARATTERIZZAZIONE ED APPLICAZIONI DEI LANGMUIR-BLODGETT FILMS - Marco Amioti	pag. 1
	LA CATODOLUMINESCENZA - Vittorio Bellani	pag. 35
	CORRELAZIONI ELETTRONICHE IN OSSIDI DI METALLI DI TRANSIZIONE - Luigi Sangaletti	pag. 63
Volume VIII - n. 1 15/1/1993	TEORIA DELLE STRINGHE IN DIMENSIONE NON CRITICA - Alberto Vancheri	pag. 1
	ROTTURA ESPlicita E SPONTANEA DI SIMMETRIE CONTINUE GLOBALI NEL MODELLO STANDARD - Antonio Defendi	pag. 25
	APPLICAZIONI DELLA $\mu^+\text{SR}$ NELLA STRUTTURA DELLA MATERIA - P. Carretta	pag. 39
	EFFETTI FOTORIFRAATTIVI IN CRISTALLI IONICI - Enrico Giulotto	pag. 75
Volume VIII - n. 2 15/4/1993	L'UNITA' DELLA SCIENZA. IL CASO DELLA FISICA. OGGI - Giorgio Salvini	pag. 1
	LA RICERCA DEL QUARK TOP: STATO ATTUALE E PROSPETTIVE FUTURE - Elisabetta Pennacchio	pag. 37
	MODELLO A TETRAEDRI PER LA FUNZIONE DIELETTRICA DI SOLIDI AMORFI - Adele Sassella	pag. 73
	INTRODUZIONE ALLE RETI NEURALI - Chiara Macchiavello	pag. 93