Identificazione dei Parametri Caratteristici di un Plasma Circolare
Tramite Rete Neuronale

Descrizione

Il presente lavoro, facente segiuto a quanto descritto precedentemente, ha il fine di:

1) introdurre altri tipi di funzioni sigmoidali;

2) produrre un set di pesi iniziali che, sempre nell'ambito della casualità, generino delle uscite nel range di quelle attese senza entrare nella saturazione della funzione di attivazione scelta;

3) determinare le costanti della funzione lineare di uscita del tipo **g(x) = cost1 · x + cost2**, dove **cost1** e **cost2** sono diverse per ogni uscita;

4) introdurre un termine di momento nella procedura di training del tipo:

$$W_{pq}(t+1) = -\mu \frac{\delta E}{\delta W_{pq}} + \alpha \cdot W_{pq}(t)$$

Per il punto 1), la funzione di attivazione scelta è:

$$g(x) = \frac{2}{1 + e^{-2\beta x}} - 1$$

che risulta compresa tra -1 ed 1 ed è derivata dalla funzione Glauber compresa, invece, tra 0 ed 1.

La derivata prima di g(x) è:

$$g'(x) = \beta \cdot \left(- g(x)^2 \right)$$

(Per β=1 si ha g(x)=tanh(x), per β=0.5 si ha g(x)=tanh(w/2))

Gli andamenti di g(x) per vari β sono riportati in Fig. 1.

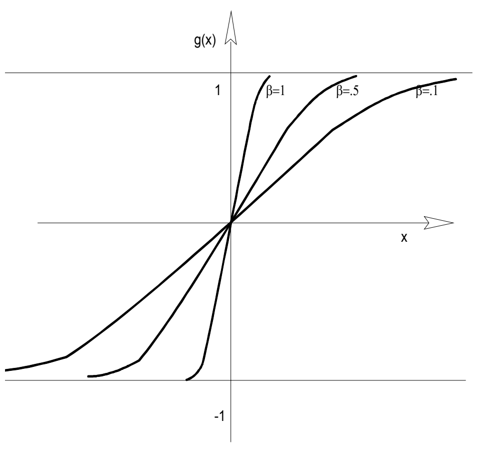


Fig. 1

Nel nostro caso si è preso **β=0.5**.

Per i punti 2) e 3) si riporta nua descrizione dettagliata della procedura utilizzata in Allegato1.

Per il punto 4) si deve dire che il termine **$\alpha \cdot W_{pq}(t)$** ha un compito di tipo "inerziale" riducendo le oscillazioni dei W(t+1) ed agevolando il passo in situazioni di costanza di andamento.

Nel nostro caso si è preso **α = 0.5**.

Con tali modifiche e/o implementazioni si sono ottenuti i risultati riportati nelle Fig. 2 e 3.

Analisi dei risultati

Come si può vedere dalla Fig.2 le distribuzioni delle uscite ottenute sono sostanzialmente identiche a quelle delle uscite attese sia come valore medio che come deviazione standard; si può notare un piccolo eccesso dei valori estremali.

Le distribuzioni degli errori sono molto più ristrette di quelle ottenute nel caso precedentemente analizzato e l'errore rms assoluto è meno della metà dell'altro e pari a rms totale = 0.131.

Sulle singole uscite abbiamo:

Valori attuali	Valori precedenti
R: rms = 0.072	0.140
Z: rms = 0.044	0.169
Λ: rms = 0.101	0.179

Gli errori percentuali sulla media, riferiti al valore massimo, sono in tutti i casi nulli in quanto si è inserita nella fase di Learning un aggiornamento della cost2 delle funzioni lineari di uscita.

Dalla Fig. 3 si possono vedere i confronti tra le uscite attese e le ottenute con evidente miglioramento rispetto alle precedenti ed, in particolare, della Z.

In Allegato 2 sono riportati i parametri significativi dell'esempio in questione.

Conclusioni

Risulta evidente la grande importanza che ha la determinazione dei pesi iniziali sulla bontà del risultato finale.

Si pensa che tale punto debba ancora essere approfondito al fine di arrivare ad un set di pesi ulteriormente più vicini a quelli che saranno i pesi definitivi ottenuti nella fase di training.

Altro fattore importante è quello della determinazione delle costanti delle funzioni di uscita differenziate: ciò ha portato a miglioramenti sia come dispersione degli errori che come centraggio del valore medio.

In ultimo, l'introduzione nel training, del termine di momento ha portato un contributo, anche se non particolarmente sensibile.

ed ottimizzata all'interno della stessa procedura di training.

Ambedue le posizioni sono, peraltro, suggerite da alcuni autori [Carter, 1987; Jacobs, 1988 et al.].

Una efficiente procedura per la determinazione del learning rate è, inoltre, fondamentale per l'ottenimento degli stessi risultati utilizzando un numero minimale di Patterns.

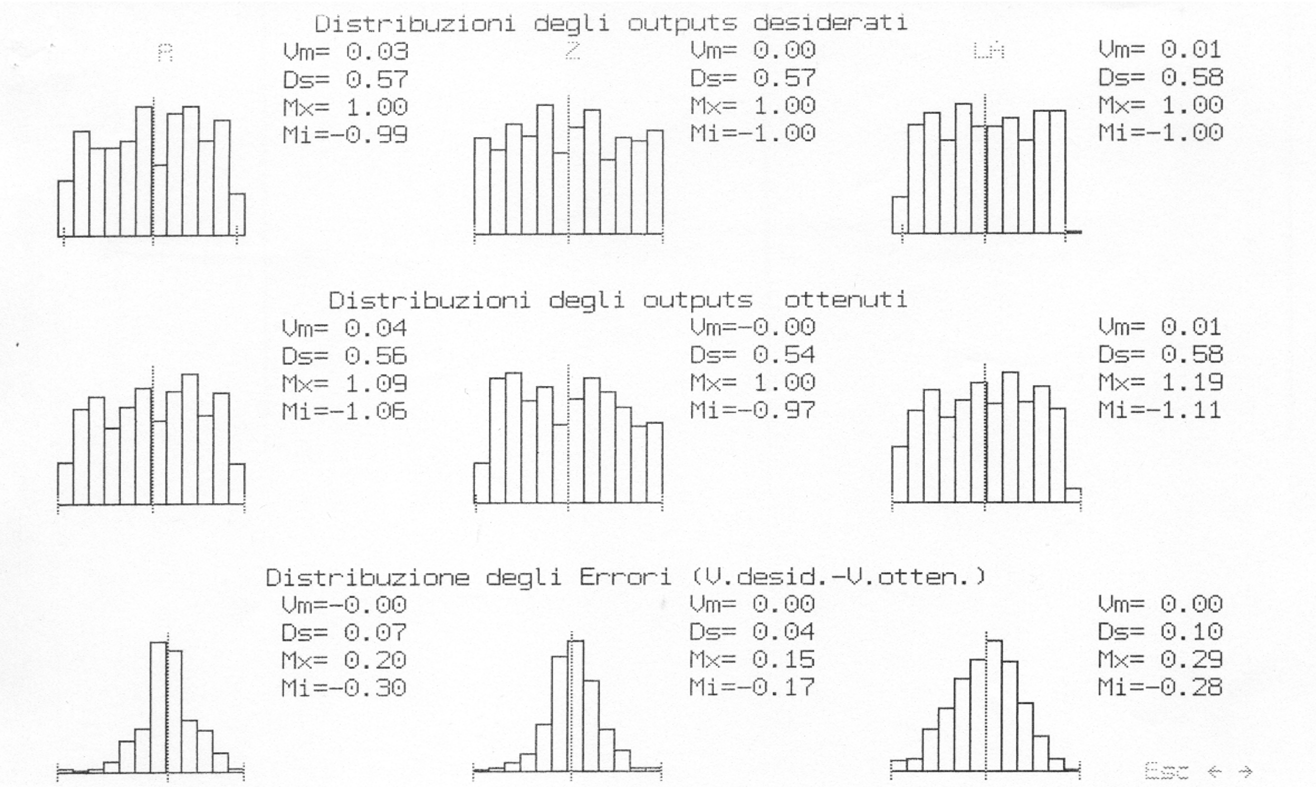


Fig. 2

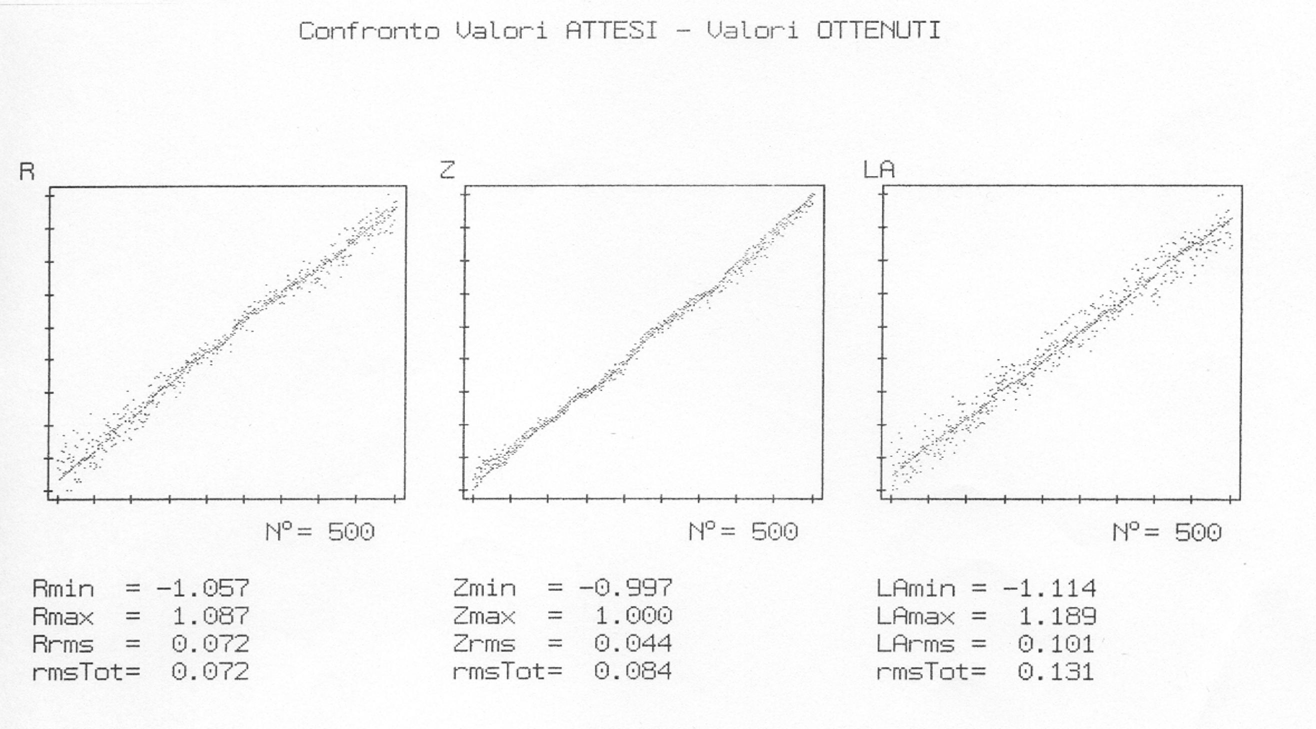
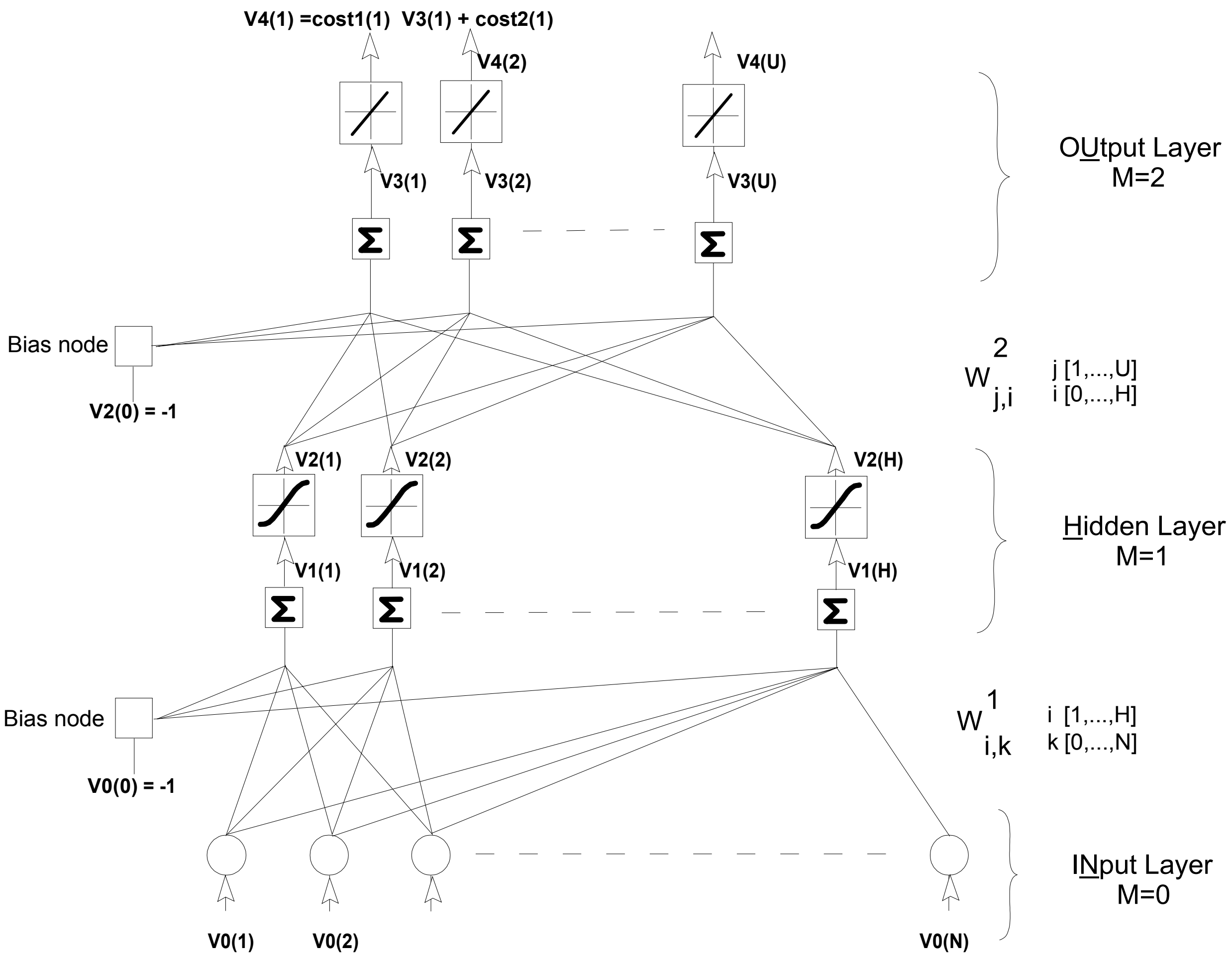


Fig. 3

Allegato 1

Procedura per il calcolo dei pesi iniziali.

Tale procedura ha il fine di ottenere, per qualsiasi vettore di Input della banca dati e con l'architettura della rete fissata, delle uscite casuali comprese tra due limiti (in questo caso -1 e +1).



Si parte dall'assunto che tutti i pesi del livello 1(tra il piano 1 e 0) siano uguali e pari a $W1_{ini}$.

Si potrà scrivere:

$$-R \leq W1_{ini} \cdot \Sigma_{k=0}^{N} V0_k \leq R$$

dove R e -R sono i valori massimio e minimo che si desidera avere in ingresso alla funzione di attivazione sigmiodale.

Risulta quindi:

$$-\frac{R}{\Sigma_{k=0}^{N} V0_k} \leq W1_{ini} \leq \frac{R}{\Sigma_{k=0}^{N} V0_k} \qquad (1)$$

effettuando la (1) tante volte quanti sono i patterns utilizzati nella fase di training (P), si otterranno P valori diversi del range per $W1_{ini}$: quello più ristretto soddisferà tutti gli altri.

Si possono ora generare i pesi del livello 1 in modo casuale tra i limiti minimi trovati: i pesi sono chiamati W1 (i,k) con $i \in [1,...,H]$ e $k \in [0,...,N]$.

Ad ogni pattern il vettore di ingresso al piano Hidden sarà:

$$\overline{V1}_i = \sum_{k=0}^{N} W1(i,k) \cdot \overline{V0}_k \qquad i \in [1,...,H]$$

le cui componenti saranno comprese tra -R e R.

La casualità può portare, però, a combinazioni tali da avere una o più componenti che, attraverso il training, rimangono costantemente positive o negative (ciò significa che su alcuni neuroni hidden possono trovarsi sempre valori dello stesso segno).

Per eliminare queste anomalie si rigenerano, sempre all'interno del range prima definito, soltanto i pesi W1 che fanno capo a quegli elementi, fino a quando non si sia raggiunto per tutte le componenti e per tutti i patterns un adeguato livello di omogeneità (si è presa una percentuale di valori positivi compresa tra il 40% ed il 60%).

Ancora per ragioni di casualità, è possibile che le varie componenti attraverso tutti i patterns occupino soltanto una parte del range richiesto (-R,R): sarà allora necessario rigenerare tutti i pesi W1(i,k) con margini più ampi fino ad ottenere un range occupato vicino a quello voluto.

Ora si devono sottoporre i P vettori V1 alla compressione effettuata dalla funzione di attivazione scelta ottenendo i P vettori V2, le cui componenti saranno, comunque, comprese tra -1 e +1.

Si utilizzano ora i vettori V2 per la determinazione dei pesi W2(j,i) seguendo la stessa procedura vista per i pesi W1(i,k) con la sola differenza che il range disiderato in uscita è ora compreso tra -1 e +1.

Ad ogni pattern, il vettore di ingresso al piano di Output sarà:

$$\overline{V3}_j = \sum_{i=0}^{H} W2(j,i) \cdot \overline{V2}_i \qquad j \in [1,...,U]$$

le cui componenti risulteranno comprese tra -1 e +1.

L'ultima fase è la trasformazione lineare dei vettori V3 in V4, che costituiscono l'uscita del piano di Output.

La trasformazione è del tipo :

$$\overline{V4}_j = \cos t1(j) \cdot \overline{V3}_j + \cos t2(j) \qquad j \in [1,...U]$$

dove le cost1(j) e cost2(j) sono calcolate in modo da avere il range di variazione ed il valore medio delle singole uscite coincidente con gli stessi parametri delle uscite desiderate. Tali costanti verranno utilizzate in seguito nella fase di training.

Allegato 2

W1min=-1.81214 W1max= 1.81214
Yes
Yes
Yes
Yes
Yes
Yes
Yes
Yes
12 -3.3052 3.6366
Altra Amp <S/N>n
Amp= 1.5

W2min=-0.49091 W2max= 0.49091
Yes
Yes
Yes
2 -0.9996 0.9024
Altra Amp <S/N>n
Amp= 1

LC(1)= 1.1299	LCC(1)= 0.1486
LC(2)= 1.2156	LCC(2)=-0.0970
LC(3)= 1.3773	LCC(3)=-0.1483

Training effettuato su 1000 casi
Testing effettuato su 500 casi
Learning rate = 0.25
Momento = 0.5

LC(i) ed LCC(i) sono le costanti delle funzioni lineari di uscita calcolate alla fine della procedura di determinazione dei pesi iniziaali.