

La funzione d'errore, infatti, può essere considerata come una funzione dell'energia del sistema: allo stato di minima energia corrisponde lo stato di equilibrio dello stesso. La funzione dell'errore, quando la funzione di trasferimento (o di attivazione) è lineare è rappresentata da una superficie a forma di iperparaboloide dotata di un minimo assoluto. L'apprendimento corrisponde al raggiungimento della configurazione di minima energia ed è garantito, in quanto l'iperparaboloide è dotata di un solo minimo assoluto. Quando la funzione di attivazione è non lineare, ma monotona come la funzione sigmoide, essa è rappresentata da un'iperparaboloide deformata, caratterizzata dalla presenza di minimi relativi. Esiste quindi il pericolo che, durante la fase di apprendimento si rimanga intrappolati in un minimo relativo assestandosi in una configurazione che impedisce il pieno sviluppo delle capacità neuronali.

4.2. Proprietà delle reti neurali

Le reti neurali si distinguono particolarmente, in quanto dotate della proprietà di **approssimazione universale**, cioè della capacità di approssimare la legge che descrive un fenomeno. I teoremi, frutto della ricerca di Hornik, Stinchcombe e White (Hornik K., 1991) stabiliscono che reti neurali *feed-forward* con un solo strato nascosto possono approssimare qualunque funzione continua, indipendentemente dalla funzione di attivazione e dalla dimensione r dello spazio degli input.

I teoremi appena citati sono importanti soprattutto dal punto di vista della implementazione; tali teoremi, infatti, ci permettono di codificare reti neurali relativamente semplici e quindi di ridurre i tempi di stima del modello.

Una ulteriore proprietà delle reti neurali è quella della **generalizzazione** ovvero la capacità della rete neurale di effettuare la previsione su dati diversi da quelli utilizzati per il suo addestramento.

Quest'ultima è un'importante caratteristica, in quanto se la rete neurale si limitasse a fornire risposte corrette Y ai soli input X degli esempi (X, Y) utilizzati nell'addestramento, essa si limiterebbe ad "apprendere a memoria" (Cammarata, 1990). Il vero apprendimento, invece, dovrebbe offrire alla rete capacità previsionali, consentendo quindi risposte Y sostanzialmente corrette anche agli input X non compresi nel *training set* e appartenenti generalmente ad un opportuno *validation set*. La probabilità che una rete neurale *feed-*

forward apprenda a calcolare un numero m di esempi di un *training set* cresce all'aumentare del numero delle connessioni (e quindi del numero di neuroni nascosti). Tuttavia, all'aumentare della complessità di una rete neurale diminuisce la sua capacità di generalizzazione e può dare seri problemi di *overfitting* (semplice memorizzazione del dato). Per aumentare la capacità di generalizzazione è allora preferibile diminuire la complessità della rete, fino al limite della funzione d'interpolazione necessaria.

4.3. Pregi e difetti delle reti neurali e loro applicazioni

Come in ogni algoritmo anche le reti neurali del tipo descritto sono dotate di pregi e difetti.

4.3.1. I punti forti

Le reti neurali si propongono come modelli induttivi in grado di superare alcuni dei limiti tipici del tradizionale approccio deduttivo. Tale approccio si basa sulla formulazione di ipotesi a priori sui sistemi oggetto d'indagine, che spesso risultano riduttive della realtà perché definiscono relazioni semplici tra le variabili in esame. L'approccio induttivo, invece, non è vincolato da ipotesi e perciò non preclude la scoperta di dipendenze più complesse e nascoste anche se, talvolta, può limitare le possibilità d'indagine sui meccanismi dei processi.

Le reti neurali inoltre sono dotate di notevoli altri vantaggi.

Apprendimento. L'attrattiva fondamentale delle reti neurali deriva dalla loro capacità di apprendere "realmente", cioè non "a memoria", come previsto dalla proprietà di generalizzazione. Ciò consente, a regime, di risolvere problemi tramite l'associazione immediata della soluzione ai dati d'input senza dover sviluppare un algoritmo applicativo: cosa particolarmente interessante se l'algoritmo è complesso, o ignoto, o troppo lungo da programmare, o troppo gravoso da eseguire. L'apprendimento, inoltre, consente di estrarre da masse di dati oscuri eventuali significati reconditi e relazioni interessanti.

Adattività. Le reti neurali possono adeguarsi flessibilmente a situazioni complesse che cambiano nel tempo, tramite il semplice riaddestramento con nuovi dati.

Robustezza. La grande tolleranza a guasti ed a disfunzioni è una dote di rilievo in applicazioni di vario tipo, quando il sistema debba operare in modo autonomo, per tempi lunghi o indefiniti, senza la possibilità di riparazioni o di controlli anche remoti, in ambienti ostili o difficili.

Velocità di calcolo. L'associazione input-output è molto rapida (poiché i calcoli sono semplici somme pesate e decisioni di soglia), anche se l'addestramento a svolgere un determinato compito può richiedere tempi molto lunghi. Le reti neurali possono costituire una valida alternativa all'esecuzione di calcoli complessi, particolarmente quando:

- i tempi di esecuzione sono critici;
- è preponderante l'esigenza di adattarsi flessibilmente a situazioni del mondo reale mutevoli o ignote o difficilmente rappresentabili in forma matematica.

4.3.2. I punti deboli

Naturalmente anche le reti neurali hanno le loro caratteristiche di debolezza, alcune delle quali vengono elencate di seguito.

Necessità di una grande casistica di esempi. Nel caso di reti ad apprendimento con supervisione, è necessaria una notevole e significativa quantità di esempi noti da sottoporre alla rete. Ciò non è sempre disponibile o facile da raccogliere o, se disponibile, non sempre si dispone di documentazione omogenea, mancando qualche elemento dei vettori che costituiscono gli esempi.

Mancanza di una teoria completa. Mancano metodi sistematici per la progettazione delle reti e criteri rigorosi per organizzare le informazioni da manipolare.

Assenza di modularità. L'analisi delle prestazioni e l'individuazione di eventuali anomalie sono rese particolarmente difficili dal fatto che ogni parte della rete contribuisce ad ogni parte della soluzione del problema.

4.3.3. I campi d'applicazione

Le reti neurali possono trovare applicazione, oltre che nell'ambito dell'analisi statistica (interpolare funzioni, completare serie storiche dotate di dati mancanti,

individuare correlazioni tra variabili sconosciute) nei campi più diversi (Cammarata, 1990):

- sistemi di supporto alle decisioni;
- diagnosi mediche;⁴
- previsione;
- visione e riconoscimento di forme grafiche;
- interpretazione di segnali;
- robotica e veicoli autonomi;
- riconoscimento della voce;
- problemi di ottimizzazione;
- modellazione e controllo di processi industriali.

L'elenco è solo parziale e sicuramente in continuo sviluppo.

4.4. Metodi di valutazione dell'efficacia dei modelli elaborati con reti neurali

La valutazione delle prestazioni è un aspetto fondamentale nel processo di sviluppo di un modello e, ciò nonostante, ha ricevuto relativamente scarsa attenzione nell'ambito della letteratura scientifica. Il problema è, inoltre, notevolmente amplificato dal mancato accordo tra gli scienziati nella valutazione delle misure più adatte a determinare l'accuratezza previsionale dei modelli (Wilmott et Al., 1985). Il primo trattato su questo argomento, infatti, è piuttosto recente e dà il via ad un'ampia panoramica di trattazioni sull'argomento con il fine di documentarne e completarne le argomentazioni (Fox, 1981).

Si pone rilievo sul fatto che il giudizio sull'accuratezza delle previsioni deve essere effettuato mediante confronti tra i valori previsti dal modello ed i corrispondenti valori osservati.

⁴ Una nuova metodologia è stata applicata per diminuire dal 30% al 3% l'errore di diagnosi medica del tumore al collo dell'utero. Tale metodologia si basa sull'utilizzo di reti neurali, chiamate Pap net, per il riconoscimento di cellule tumorali nell'esame ginecologico del Pap test (Cozza, 1997).

Un primo indice quantitativo che fornisce una misura dell'associazione (correlazione) tra i valori previsti (p_j) e quelli osservati (o_j) è il coefficiente di correlazione di Pearson (r).

Talvolta, a quest'ultimo, viene aggiunto il suo quadrato (r^2), il coefficiente di determinazione (o "varianza spiegata").

I coefficienti r ed r^2 danno un'indicazione sul legame esistente tra le osservazioni ed i valori previsti, ma non bastano ai fini di una valutazione della prestazione di un modello. La loro grandezza, infatti, non è consistentemente relazionata all'accuratezza della previsione, soprattutto quando l'accuratezza sia definita come il grado con cui le osservazioni previste si avvicinano ai valori delle corrispondenti osservate. Willmott dimostra (Willmott, 1982) che la correlazione tra variabili previste molto diverse dalle corrispondenti osservate può avvicinarsi molto al valore massimo (1.0). Valori alti di r ed r^2 possono, quindi, essere ingannevoli, in quanto non correlati con l'entità delle differenze tra valori previsti e valori simulati.

Si ritiene che una stima sufficientemente significativa dell'accuratezza e della precisione delle previsioni di un modello sia data da misure realizzate sulle differenze tra i valori previsti e quelli osservati, dette anche *indici di differenza (difference measures)*, (Willmott et Al., 1985). Le formule statistiche di tali indici sono riportate in tab. 4.1.

MAE e RMSE esprimono entrambi il valore medio dell'errore commesso dalla previsione ma la differenza tra i due consiste nel fatto che MAE è meno sensibile di RMSE agli estremi delle differenze tra i valori previsti e quelli osservati. RMSE, inoltre, viene scomposto in $RMSE_s$ ed $RMSE_u$ in modo da dare una valutazione di quanto le previsioni realizzate dal modello siano affette da errori di natura sistematica. Affinché una previsione possa ritenersi "buona", $RMSE_s$ dovrebbe assumere valori vicini allo zero, mentre $RMSE_u$ dovrebbe avvicinarsi a RMSE. Gli indici di Willmott forniscono un'indicazione dell'errore relativo medio.

Oltre a statistiche atte alla descrizione dello scarto tra le due variabili P ed O è importante riportare anche statistiche che descrivano l'andamento di tali variabili, come le loro medie, le loro deviazioni standard, l'intercetta (a) ed il coefficiente angolare (b) della retta di regressione ai minimi quadrati.

Tabella 4.1. Statistiche per la valutazione della precisione delle previsioni realizzate con modelli neurali.

| Parametri statistici per la valutazione delle capacità previsionali delle reti neurali | |
|---|--|
| Coefficiente di correlazione | $R = \frac{\sum_{i=1}^N (P_i - \bar{P})(O_i - \bar{O})}{\left[\sum_{i=1}^N (P_i - \bar{P})^2 \cdot \sum_{i=1}^N (O_i - \bar{O})^2 \right]^{1/2}}$ |
| Regressione lineare ⁵ | $\bar{P}_i = a + bO_i$ |
| Errore Medio Assoluto (Mean Absolute Error) | $MAE = N^{-1} \sum_{i=1}^N P_i - O_i $ |
| Radice dell'errore quadratico medio ⁶ (Root Mean Square Error) | $RMSE = \left[N^{-1} \sum_{i=1}^N (P_i - O_i)^2 \right]^{1/2}$ |
| Errore quadratico medio sistematico (Systematic Mean Square Error) | $RMSE_s = \left[N^{-1} \sum_{i=1}^N (\bar{P}_i - O_i)^2 \right]^{1/2}$ |
| Errore quadratico medio casuale (Unsystematic Mean Square Error) | $RMSE_u = \left[N^{-1} \sum_{i=1}^N (P_i - \bar{P}_i)^2 \right]^{1/2}$ |
| Indice di Willmott, d ₁ (willmott Index d ₁ of agreement) | $d_1 = 1 - \left[\frac{\sum_{i=1}^N P_i - O_i }{\sum_{i=1}^N (P_i + O_i)} \right], \quad 0 \leq d_1 \leq 1$ |
| Indice di Willmott, d ₂ (Willmott Index d ₂ of agreement) | $d_2 = 1 - \left[\frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (P_i + O_i)^2} \right], \quad 0 \leq d_2 \leq 1$ |

⁵ In questa formula a e b sono l'intercetta e il coefficiente angolare della retta di regressione ai minimi quadrati.

⁶ $RMSE^2 = RMSE_s^2 + RMSE_u^2$.

dove N è il numero di esempi, O_i sono i valori osservati, \bar{O} è la media dei valori osservati, P_i sono i valori previsti dal modello, \bar{P} è la media dei valori previsti dal modello, $P_i' = P_i - \bar{O}$, $O_i' = O_i - \bar{O}$.

La valutazione delle capacità previsionali di un modello prevede anche dei test di previsione di eventi acuti, cioè di eventi che superano determinati valori di soglia.

I test presentati fino ad ora riportano valori medi: incapaci di “dirci” quanto le previsioni di eventi acuti si discostano da questi ultimi e quindi di darci informazioni sulla reattività della rete ai fenomeni di superamento delle soglie.

La rete neuronale, infatti, deve essere in grado non solo di fornire risultati mediamente precisi, ma anche di realizzare previsioni aderenti alla realtà in momenti critici.

A tal fine si aggiungono dei test di verifica delle previsioni di eventi acuti utilizzando (Ryan W. F., 1995):

- tabelle standard di contingenza;
- analisi statistiche basate sulle tabelle di contingenza.

Le tabelle di contingenza riportano, per entrambe le variabili P ed O del test set, il numero di eventi favorevoli e quello degli sfavorevoli al verificarsi di un dato evento, che nel nostro caso è il superamento di una soglia d'inquinamento atmosferico.

In tabella 4.2. si riporta un esempio di tabella di contingenza per il superamento del livello di attenzione dell'ozono (DM 25/11/96), mentre in tabella 4.3. si riportano le statistiche utilizzate in qualità di test di verifica della previsione dell'evento a cui si riferisce la tabella di contingenza.

Tabella 4.2. *Tabella di contingenza per il superamento della soglia d'attenzione dell' O_3 .*

| Tabella di contingenza per $O_3 > 180 \text{ mg/m}^3$ | | |
|--|--|---|
| | Eventi favorevoli per i valori previsti | Eventi sfavorevoli per i valori previsti |
| Eventi favorevoli per i valori osservati | (A) | (B) |

| | | |
|--|-----|-----|
| Eventi sfavorevoli per i valori osservati | (C) | (D) |
|--|-----|-----|

Tabella 4.3. Statistiche applicate alla tabella di contingenza.

| Statistiche di contingenza per O₃ > 180 mg/m³ | |
|---|---|
| <i>POD</i> (Probability of detection) | $POD = \frac{A}{(A + B)}$ |
| <i>MISS</i> (Miss Rate) | $MISS = 1 - POD = \frac{B}{(A + B)}$ |
| <i>FAR</i> (False Alarm Rate) | $FAR = \frac{C}{(C + A)}$ |
| <i>CNULL</i> (Correct Null Forecast) | $CNULL = \frac{D}{(D + C)}$ |
| <i>CSI</i> (Critical Success Index) | $CSI = \frac{A}{(A + B + C)}$ |
| <i>TSS</i> (True Skill Scores) | $TSS = POD + CNULL - 1$ |
| <i>S</i> (Heidke Skill Score) | $S = \frac{2(AD - BC)}{[B^2 + C^2 + 2AD + (B + C)(A + D)]}$ |

Le Statistiche di verifica delle previsioni hanno i seguenti significati: il [POD] misura la percentuale degli eventi d'inquinamento che sono stati correttamente previsti; il [MISS] riporta la percentuale degli eventi d'inquinamento verificatisi ma non previsti; la variabile [FAR] misura la tendenza del modello a prevedere eventi di superamento della soglia quando in realtà non si sono verificati; una misura dell'abilità del modello nel prevedere eventi "reali" di non inquinamento è data da [CNULL]; l'indice [CSI] riporta una percentuale degli avvenimenti di previsione che non hanno avuto successo mentre [TSS] riporta il punteggio degli eventi previsti con successo; [S] infine è una misura dell'abilità predittiva del modello data dal rapporto tra le previsioni effettuate con successo sul set di

dati e le previsioni con successo su un set di dati (della stessa numerosità del precedente) estratti a caso dalla serie iniziale.