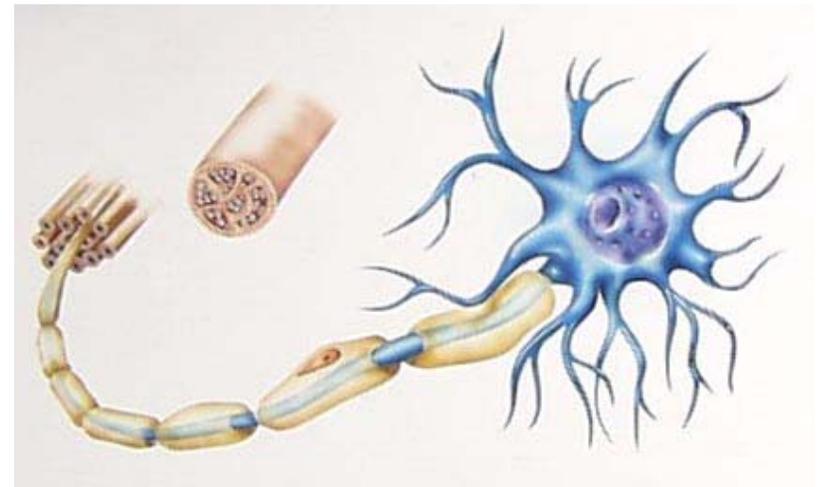
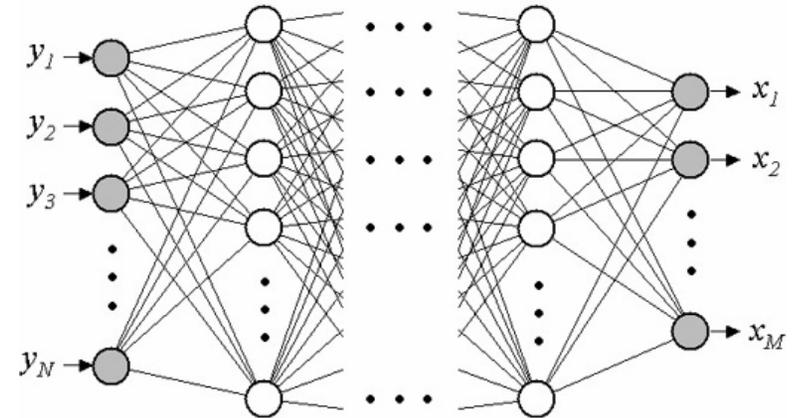
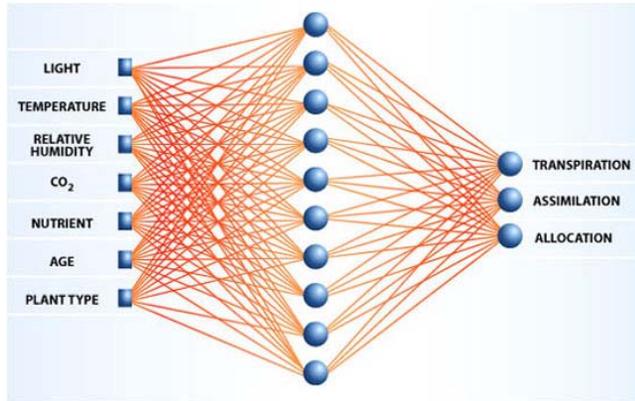
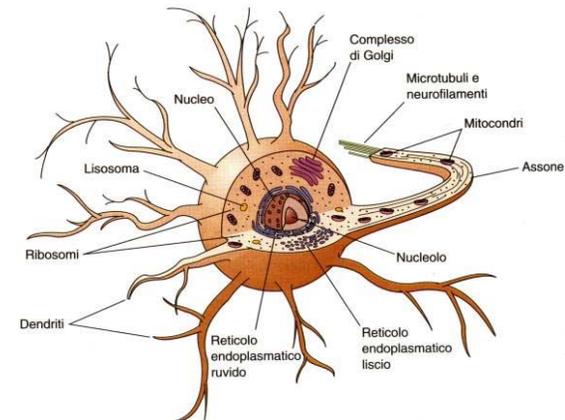
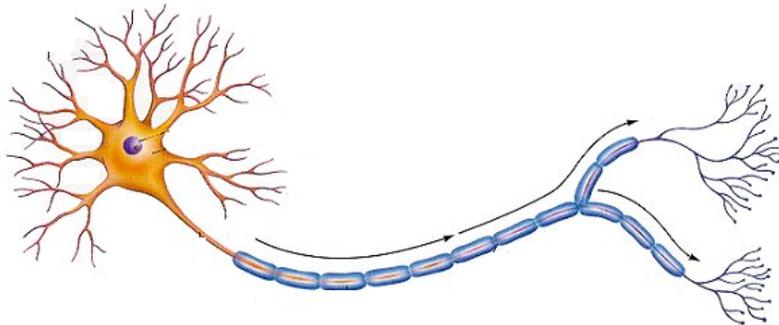
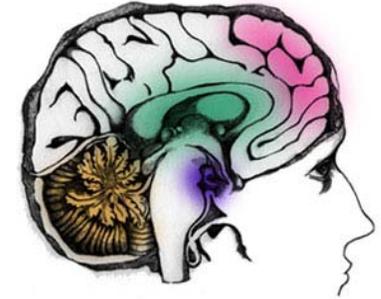


Reti Neurali Artificiali



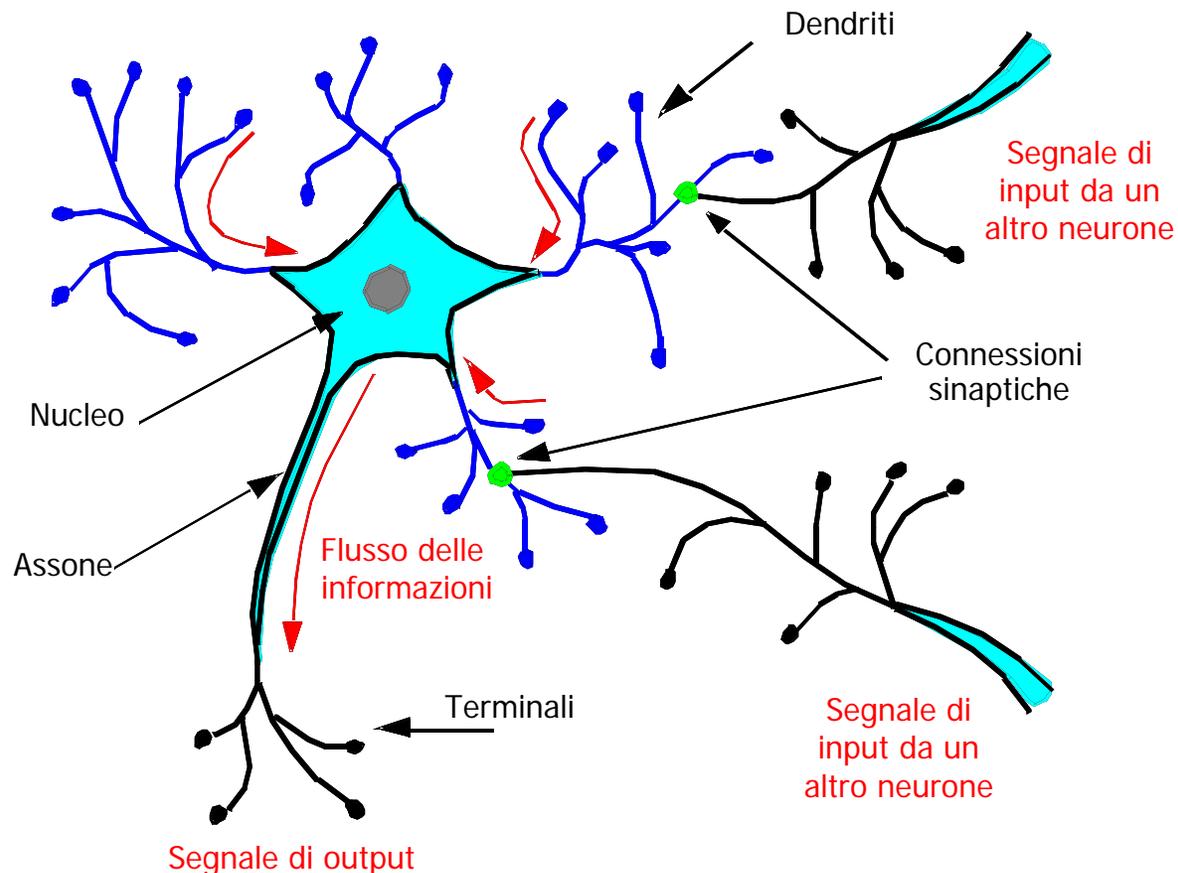
Introduzione

- **ANN**: **A**rtificial **N**eural **N**etwork
- Una rete neurale artificiale è una macchina progettata per simulare il funzionamento del **cervello umano**, implementata fisicamente utilizzando componenti elettronici o simulata tramite software su computer digitali (Haykin, 1999).
- Le reti neurali sono composte da elementi semplici (**neuroni**, nodi, unità).
- Le informazioni, i **segnali**, fluiscono tra i neuroni attraverso le connessioni.
- Le **connessioni** sono **pesate** in modo da regolare il flusso informativo.
- L'informazione, i segnali pesati, viene accumulata nei neuroni ed una **funzione di attivazione** o di trasferimento (lineare o non lineare) la trasforma nel segnale di output.

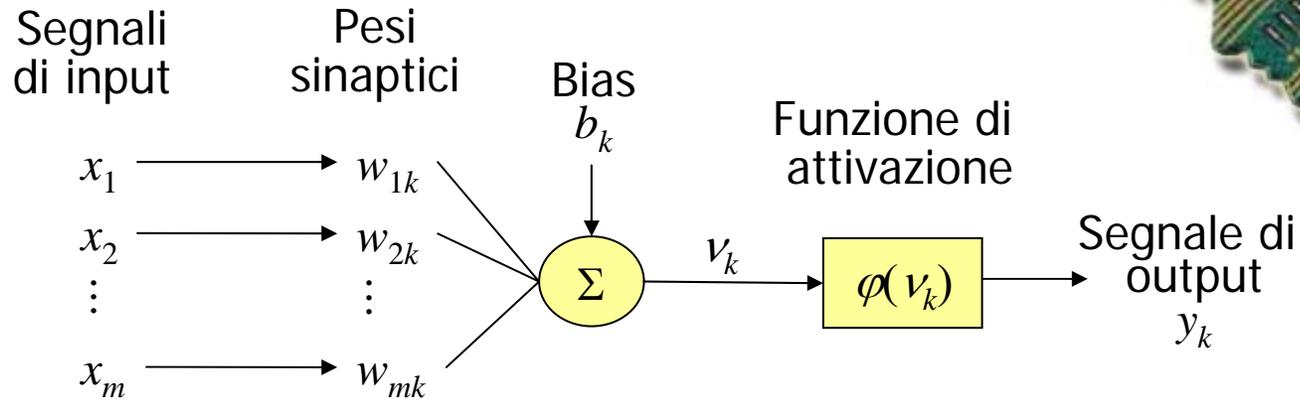
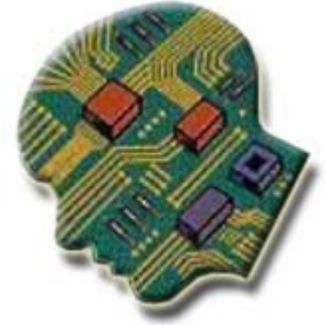


Il neurone biologico

- Il corpo centrale del neurone, il **nucleo**, somma i segnali di input provenienti dalle **sinapsi** collegate ai **dendriti** di altri neuroni. Quando il segnale raggiunge una soglia limite il neurone genera un segnale di output verso altri neuroni. Si dice che il neurone “fa fuoco”.



Il neurone astratto

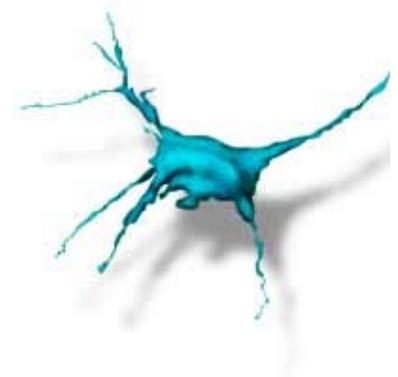


$$\mu_k = \sum_{j=1}^m w_{jk} x_j \quad \Rightarrow \quad v_k = (\mu_k + b_k) \quad \Rightarrow \quad y_k = \varphi(v_k)$$



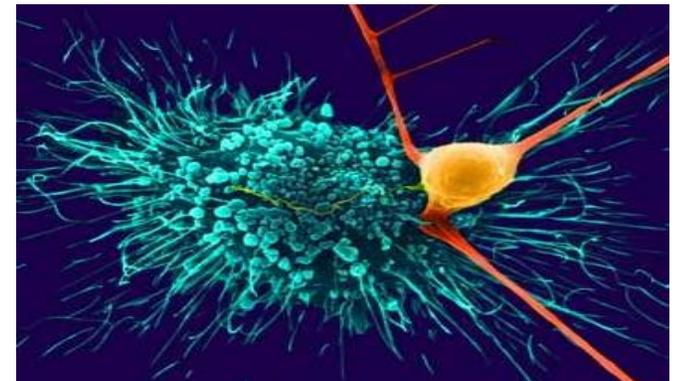
Neurone e transistor a confronto

- L'unità base del cervello umano è il neurone.
- L'unità base di un processore, CPU, è il transistor.
- Tempi caratteristici:
 - neurone $\sim 10^{-3}$ s;
 - transistor $\sim 10^{-9}$ s.
- Consumo energetico per singola operazione:
 - neurone $\sim 10^{-16}$ J;
 - transistor $\sim 10^{-6}$ J.
- Numero di unità base:
 - nel cervello $\sim 10^{10}$ neuroni;
 - in un processore $\sim 10^8 - 10^9$ transistor;
- Numero di sinapsi (interconnessioni) nel cervello umano: $\sim 60,000$ miliardi.
- Il **cervello** è assimilabile ad un complesso sistema di elaborazione delle informazioni altamente **non lineare** e **parallelizzato**.



Approfondimento

- Una rete può essere vista come un sistema in grado di dare una risposta ad una domanda o fornire un output a fronte di un dato di input.
- Il legame input \rightarrow output, ovvero la funzione di trasferimento della rete, non viene programmato ma è semplicemente ottenuto tramite un processo di apprendimento basato su dati empirici.
- In modalità supervisionata, l'algoritmo di apprendimento modifica i parametri caratteristici, pesi e bias, in modo da avvicinare la risposta della rete ai valori sperimentali forniti.
- Le principali caratteristiche della rete si avvicinano così a quelle del cervello:
 - capacità di apprendere dall'esperienza (misura di dati in campo);
 - alta elasticità di interpretazione dei dati di input (ovvero "resistenza al rumore" o "capacità di comprendere dati rumorosi");
 - discreta capacità di estrapolazione.



La storia

- La storia delle reti neurali ha inizio con **McCulloch** e **Pitts** nel **1943**. McCulloch era uno **psichiatra** e **neuro-anatomista**, mentre Pitts era un **matematico**.



La collaborazione dei due studiosi condusse alla descrizione del calcolo logico della rete neurale che unisce la neurofisiologia alla logica matematica.

- Nel **1949 Hebb** postulò la prima legge di apprendimento auto-organizzato (apprendimento senza insegnante).
- Nel **1958 Rosenblatt**, nel suo lavoro sul perceptrone, propose il primo modello di apprendimento supervisionato (apprendimento con insegnante).
- Il perceptrone realizzato da Rosenblatt sembrò porre le basi per le successive reti neurali, in grado di apprendere qualunque tipo di funzione.
- Nel **1969 Minsky** e **Papert** dimostrarono matematicamente che esistono dei limiti fondamentali alle capacità di apprendimento di un singolo neurone. Questo risultato portò gli scienziati ad abbandonare temporaneamente la ricerca relativa alle ANN.
- Negli anni **'80** le ANN tornarono alla ribalta con l'introduzione di uno o più **livelli intermedi**. Tali reti, in grado di correggere i propri errori, superarono i limiti del Perceptrone di Rosenblatt rivitalizzando la ricerca in tale settore.

Caratteristiche principali

- 1. Non linearità:** la presenza di funzioni di attivazione non lineari rende la rete intrinsecamente non lineare.
- 2. Uniformità di analisi e progettazione:** i neuroni rappresentano un ingrediente comune delle reti neurali. Ciò rende possibile la condivisione di teorie e algoritmi di apprendimento in differenti campi applicativi.
- 3. Mappa input-output:** la rete apprende attraverso esempi basati su di una sequenza di dati ingresso-uscita (apprendimento supervisionato). Nessuna assunzione a priori viene effettuata sui parametri del modello.
- 4. Adattività:** le reti neurali hanno la capacità di adattare i loro pesi a variazioni dell'ambiente circostante. In particolare una rete neurale istruita ad operare in un dato ambiente, può essere addestrata nuovamente a lavorare in un altro. Inoltre se la rete lavora in un ambiente che sta evolvendo nel tempo, questa può essere progettata per variare i propri pesi in tempo reale (elaborazione adattiva dei segnali, controllo adattivo, ...).
- 5. Parallelismo:** grazie alla propria struttura la rete può essere implementata a livello hardware per particolari scopi di calcolo richiedenti elevata velocità computazionale.

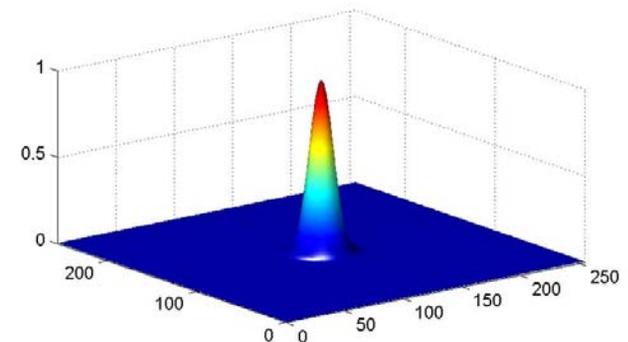
Caratteristiche principali

- 6. Tolleranza guasti:** una rete neurale implementata in forma hardware presenta virtualmente una buona tolleranza ai guasti, nel senso che le sue prestazioni degradano lentamente in presenza di malfunzionamenti. Ciò è dovuto alla forte delocalizzazione dell'informazione tra i vari neuroni, che riduce l'eventualità di accadimento di guasti catastrofici e garantisce una degradazione graduale delle prestazioni.
- 7. Analogia neuro-biologica:** la progettazione di reti neurali è motivata dalla loro analogia con il cervello umano che è la prova vivente che il calcolo in parallelo, *fault tolerant*, non è solo fisicamente possibile, ma anche veloce e potente.



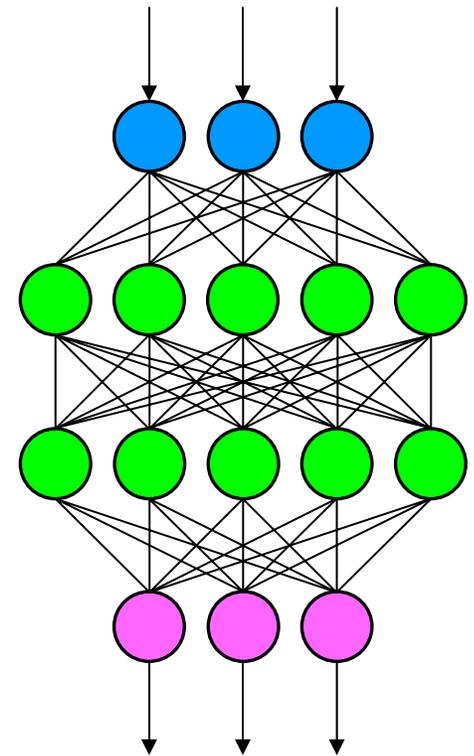
Caratterizzazione della rete

- Sono tre gli aspetti principali che descrivono e caratterizzano una ANN:
 - **architettura della rete**. Struttura e tipologia delle connessioni tra i nodi, numero di livelli intermedi;
 - **modalità di apprendimento**. Determinazione dell'intensità delle connessioni tra i nodi (calcolo dei pesi e dei bias);
 - **dipendenza funzionale** che lega l'input all'output del neurone. Tipologia della funzione di attivazione.



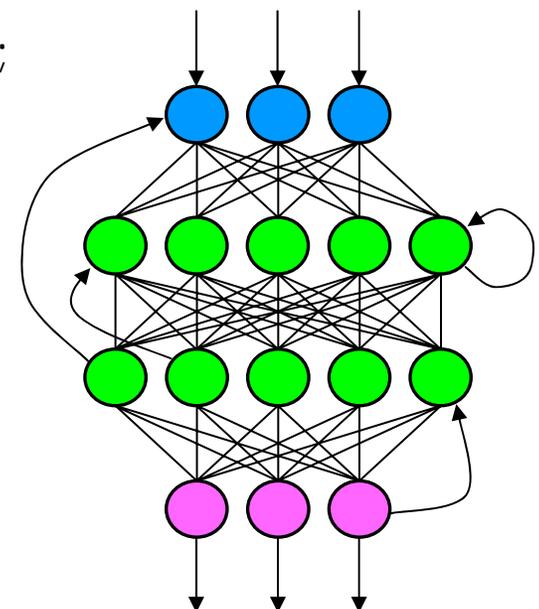
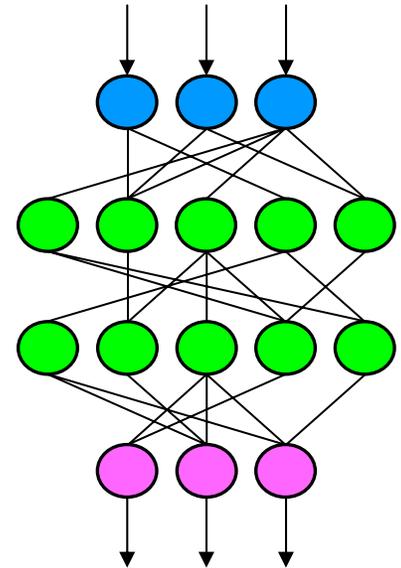
Architettura della rete

- I **nodi di input** rappresentano la sorgente dei dati. Come tali non sono considerati come un effettivo livello in quanto non contribuiscono al processamento dei dati.
- I **nodi di output** producono la risposta finale della rete.
- I nodi appartenenti ai **livelli intermedi nascosti** (se presenti), insieme ai nodi di output, trasformano il segnale di input in quello di output.
- Il numero di nodi per livello, il numero di livelli intermedi e le connessioni tra nodi definiscono l'**architettura** della rete neurale.



Architettura della rete

- Le tipologie di reti neurali sono molteplici:
 - reti prive di livelli intermedi, cioè reti input-output;
 - reti con uno o più livelli intermedi, hidden layer;
 - reti completamente connesse;
 - reti parzialmente connesse;
 - reti puramente feedforward, flusso unidirezionale;
 - reti ricorsive con informazione in feedback;
 - reti statiche;
 - reti adattive, a memoria dinamica;
 - ...



Architettura della rete

- La struttura più classica di una rete neurale prevede:
 - **un livello di input;**
 - **uno o più livelli intermedi** (hidden layer);
 - **un livello di output.**
- Ogni strato ha un numero variabile di nodi. Spesso la struttura è triangolare decrescente dall'ingresso all'uscita della rete.
- **Cybenko**, 1989, ha dimostrato che un solo livello intermedio è sufficiente per rappresentare qualsiasi funzione continua.
- **Haykin**, 1992 (1999), aggiunge che la scelta di un solo livello intermedio non sempre conduce alla migliore configurazione possibile (soprattutto nel caso di identificazione di processi altamente non lineari).



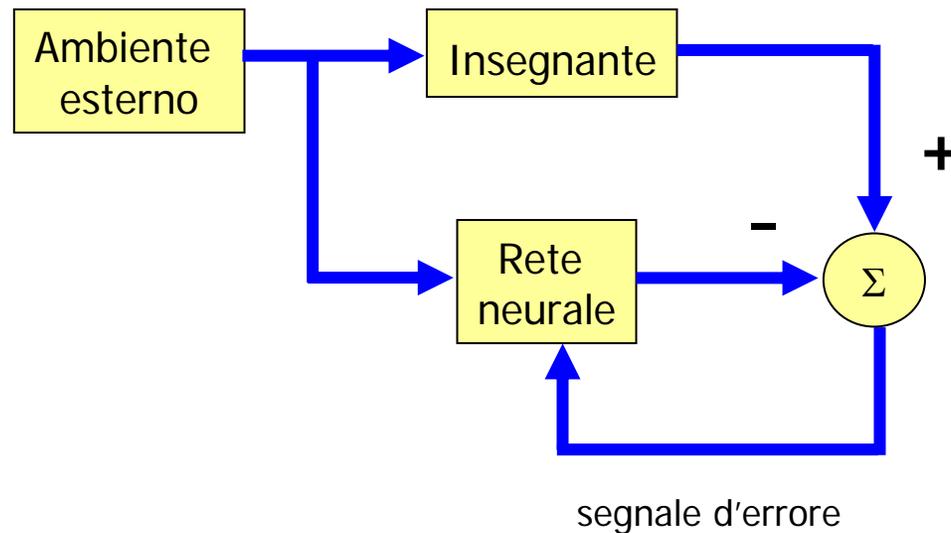
George Cybenko



Simon Haykin

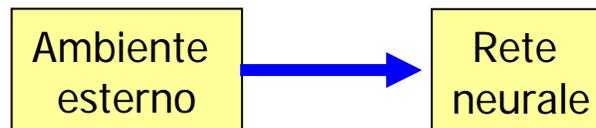
Modalità di apprendimento

- L'apprendimento della rete, che consiste nella determinazione dei pesi sinaptici ottimali, può avvenire secondo due modalità distinte:
 - **APPRENDIMENTO SUPERVISIONATO (con insegnante)**
 - Nel corso dell'apprendimento, un insegnante fornisce le risposte corrette rispetto ad un insieme di dati di input.
 - I pesi vengono cioè modificati sulla base dell'errore commesso dalla rete rispetto ai dati reali di output.



Modalità di apprendimento

- **APPRENDIMENTO NON SUPERVISIONATO (senza insegnante)**
 - Non c'è un insegnante, quindi non esiste un contributo di feedback.
 - La rete scopre da sola le correlazioni esistenti tra i dati forniti.
 - I pesi variano nel corso dell'apprendimento in base ad una **regola definita a priori** che non utilizza l'errore rispetto alla realtà.
 - Si utilizza una **misura** che è indipendente dal problema specifico.
 - Si possono utilizzare delle regole euristiche che trasformano il segnale esterno di input in "**premio**" o in "**punizione**".

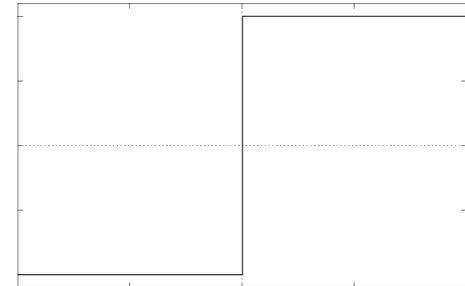


Funzione di attivazione

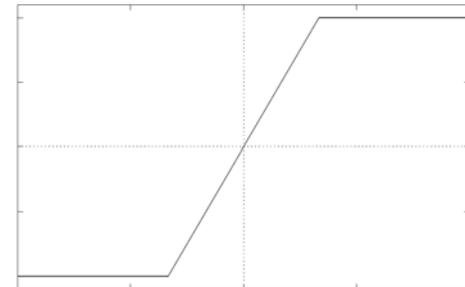
- La funzione di attivazione, $\varphi(v)$, definisce l'output di un neurone in termini di livello di attività dei suoi input.

$$\mu_k = \sum_{j=1}^m w_{jk} x_j \quad \Rightarrow \quad v_k = (\mu_k + b_k) \quad \Rightarrow \quad y_k = \varphi(v_k)$$

- Funzione soglia** $\varphi(v) = \begin{cases} 1 & v \geq 0 \\ 0 & v < 0 \end{cases}$

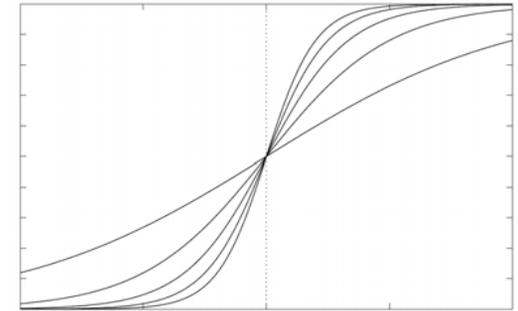


- Funzione lineare a tratti** $\varphi(v) = \begin{cases} 1 & v \geq 1/2 \\ \mu v & -1/2 \leq v \leq 1/2 \\ 0 & v \leq -1/2 \end{cases}$

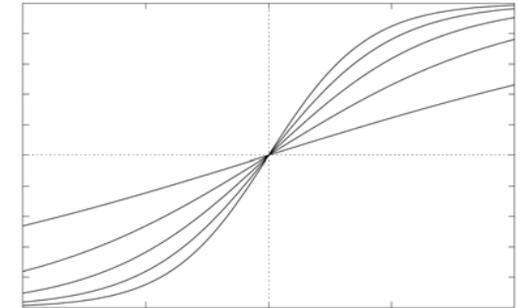


Funzione di attivazione

- **Funzione sigmoideale** $\varphi(v) = \frac{1}{1 + \exp(-\mu v)}$



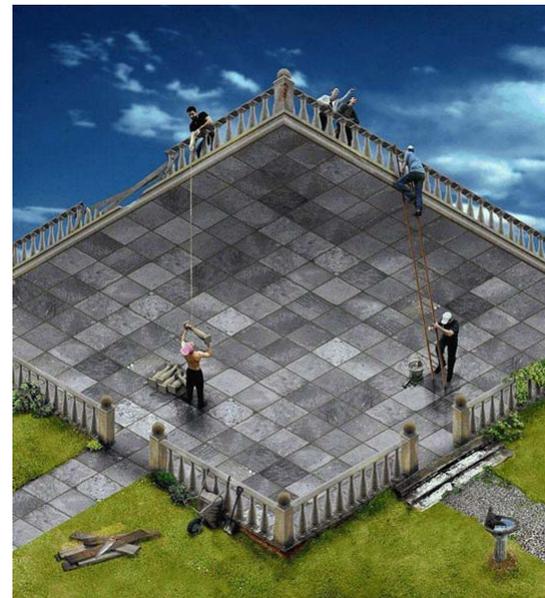
- **Funzione tangente iperbolica** $\varphi(v) = \text{Th}\left(\frac{\mu v}{2}\right) = \frac{1 - \exp(\mu v)}{1 + \exp(\mu v)}$



N.B.: le funzioni di attivazione assumono valori compresi tra 0,...1 o -1,...1. Ciò è coerente col fatto che gli input alla rete sono normalizzati al fine di mantenere limitati e controllati i segnali che viaggiano all'interno della rete.

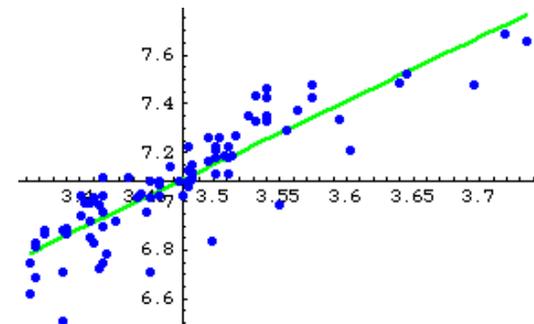
Apprendimento BP

- Tra le tecniche di apprendimento supervisionato la più nota ed utilizzata è quella definita: Error **B**ack **P**ropagation Algorithm, **BP**.
- Tale algoritmo è costituito da due fasi:
 - **forward** (in avanti);
 - **backward** (all'indietro);
- Nella fase **forward**, i pattern di apprendimento (coppie di dati input-output) sono presentati ai nodi di input. La risposta della rete si sviluppa lungo i livelli intermedi fino ad arrivare ai nodi di output. Durante questa fase i **pesi ed i bias** rimangono **invariati**.
- Nella fase **backward**, l'errore esistente tra il processo reale e la risposta della rete viene calcolato e propagato all'indietro attraverso i nodi. Tramite opportune formule di aggiornamento i **pesi ed i bias** vengono **modificati** fino al primo livello della rete (quello di input).



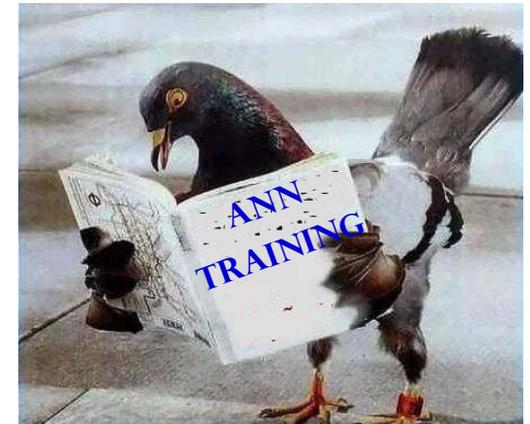
Apprendimento BP

- Per certi versi l'aggiornamento dei pesi e dei bias nel corso dell'addestramento della rete può essere visto come un problema di **regressione di parametri** ovvero di ottimizzazione di una funzione obiettivo volta a **minimizzare l'errore medio** compiuto sull'insieme dei pattern di apprendimento.
- Attenzione però al problema dell'**overfitting** (detto anche **overlearning**). Il rischio è che la rete apprenda in modo ottimo la risposta ai pattern di input-output perdendo però la capacità di generalizzazione e di risposta a dati di input non ancora sperimentati.
- È necessaria una procedura di **convalida** dell'apprendimento (**cross-validation**) basata su di un insieme di pattern diverso da quello di apprendimento.

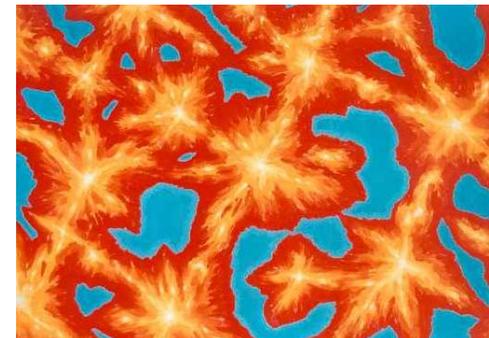


Ulteriori parametri

- La procedura di apprendimento si basa su due ulteriori parametri, entrambi limitati all'intervallo 0,...1:
 - **tasso di apprendimento**, α , (learning rate)
 - determina la velocità con cui la ANN impara: valori alti velocizzano il processo di apprendimento ma possono dare problemi di convergenza e generare instabilità.
 - **momentum factor**, β
 - considera l'evoluzione nel tempo dei pesi: la sua introduzione permette di incrementare la velocità del procedimento mantenendo elevati valori del tasso di apprendimento senza incorrere in oscillazioni e problemi di convergenza.



Un esempio



Neural network type: Back propagation with momentum factor
Activation function: Logistic sigmoid + Linear function
Learning Algorithm: History Stack Adaptation

4 ! Layers number (including input and output layers)
2 ! Neurons number at level: 1
4 ! Neurons number at level: 2
2 ! Neurons number at level: 3
1 ! Neurons number at level: 4

200 ! Total learning patterns
.50000000 ! alfa
.90000000 ! beta
.10000000 ! mu

Biases for each layer

Layer n. 2 -1.10191428411417 -.306607804280631 .396983628052738 .229701066848763E-01
Layer n. 3 1.32438158863708 .895365411961350
Layer n. 4 12.1461254510109

Weights for each layer

Layer n. 1 --> 2
-.522050637568751 .247668247408662 1.36848442775682 .784256667945155
-2.36688633725843 -2.47312321369078 1.81072640204618 .871974901958030
Layer n. 2 --> 3
1.78023653514858 1.68340336532185
-.751951327532474 .906457768573295
-1.66948863690154 -.743060896398274
-.766335553627057 -.263315123659310
Layer n. 3 --> 4
-4.61959708139503
-2.87806024443973

End of neural network

Algoritmi di identificazione

- **RICERCA LOCALE**

- Metodi del **primo ordine**

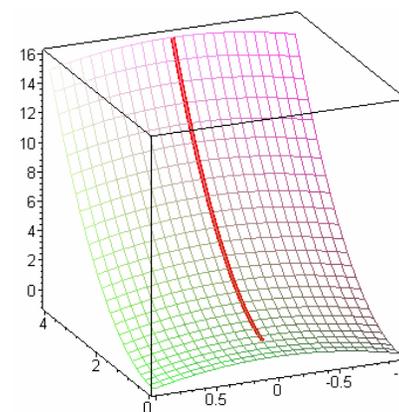
- **Gradiente** (steepest descent)

- semplice;
- possibili minimi locali;
- divergenza della ricerca.

- Metodi del **secondo ordine**

- **Levenberg-Marquardt**

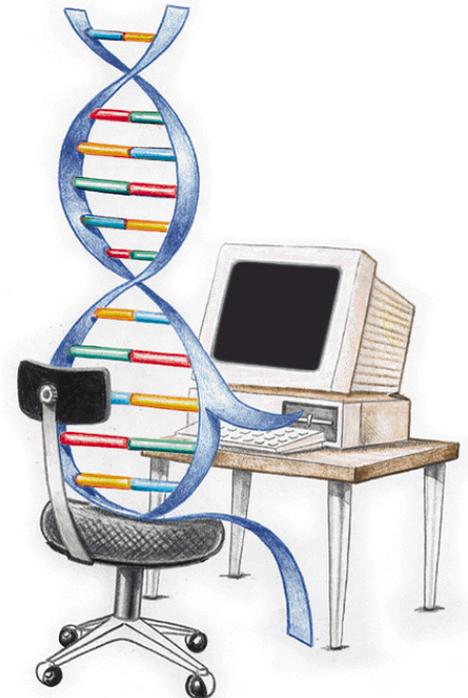
- più articolato e complesso;
- possibili minimi locali;
- maggiore occupazione di memoria;
- tempi più lunghi di apprendimento;
- minimizzazione della funzione obiettivo più spinta;



Algoritmi di identificazione

• RICERCA GLOBALE

- Simulated Annealing;
- Algoritmi genetici;
- Tabu search;
- Metodi multidimensionali specifici (metodi robusti).
- Caratteristiche:
 - implementazione complessa;
 - Costo elevato in termini di memoria e tempi di calcolo;
 - individuazione del minimo globale.



Modalità di addestramento

- Una volta raccolto l'insieme dei dati di apprendimento si procede con l'addestramento della rete.
- La presentazione di un intero insieme di pattern è detto **epoca**.
- L'addestramento è basato sulla presentazione iterativa delle epoche secondo una **sequenza casuale** dei pattern al fine di evitare apprendimenti specifici con conseguente perdita di generalità.
- **Addestramento PATTERN MODE**
 - L'aggiornamento dei pesi e dei bias avviene **a valle di ogni pattern** proposto alla rete. Per ogni pattern si effettua la procedura **forward** e quella **backward**. Se un insieme di dati di apprendimento è composto da N pattern, vengono effettuati N aggiornamenti dei pesi e dei bias per ogni epoca.
- **Addestramento BATCH MODE**
 - L'aggiornamento dei pesi e dei bias avviene **solo dopo aver presentato tutti i pattern** cioè alla fine di ogni epoca. La procedura **forward** propone N pattern. Si memorizzano gli N errori locali ed i loro gradienti locali. Alla fine se ne fa una media e si effettua la procedura **backward**.



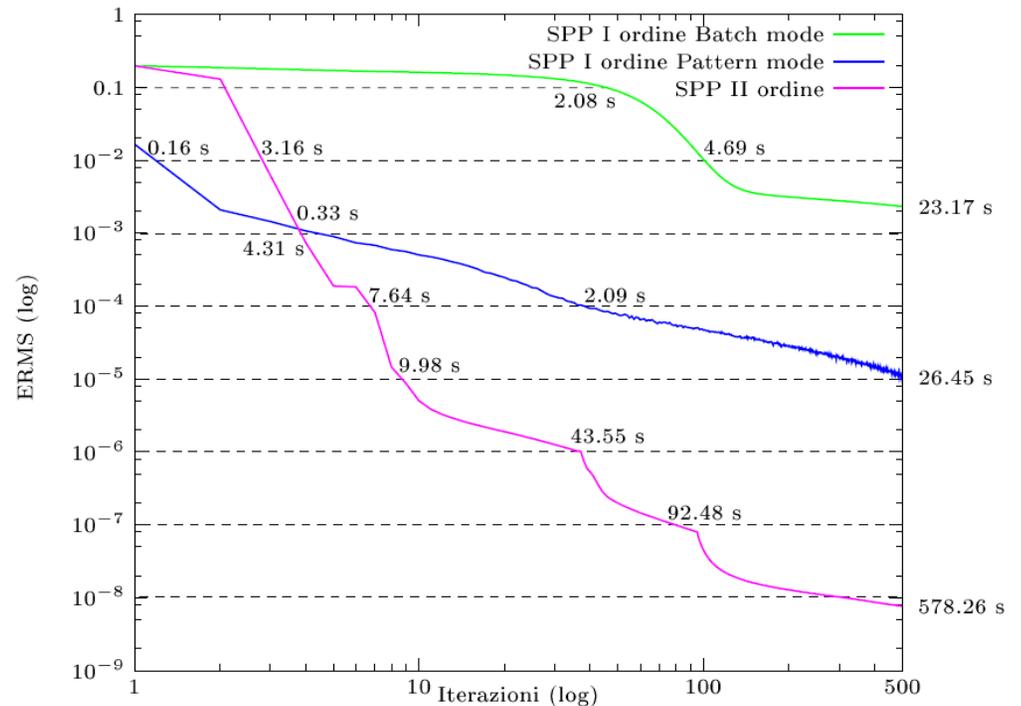
Criteri di convergenza

Non esistono dei criteri universali per valutare l'apprendimento della rete.

1. Si può focalizzare l'attenzione sul grado di identificazione relativo all'insieme dei **pattern di apprendimento**. Ad esempio controllando:

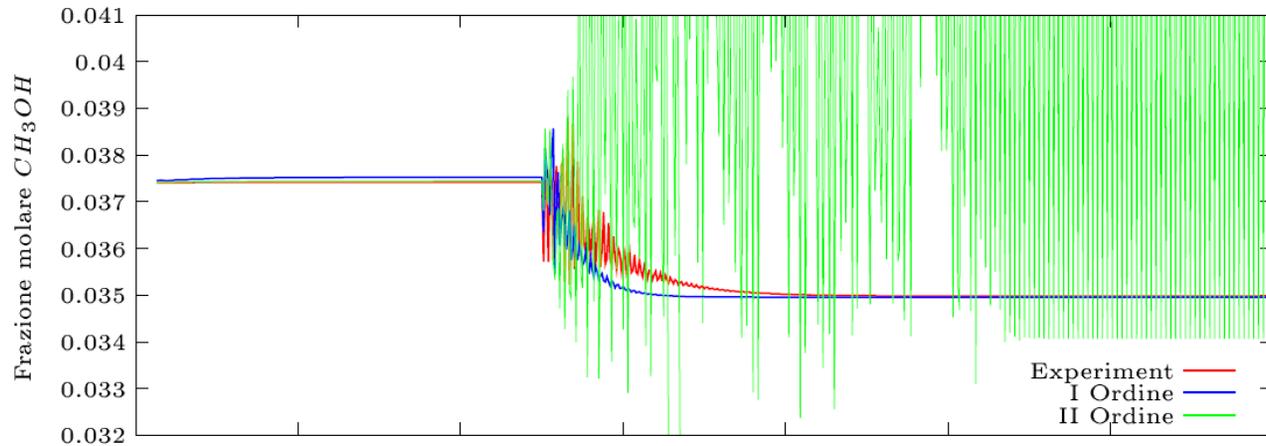
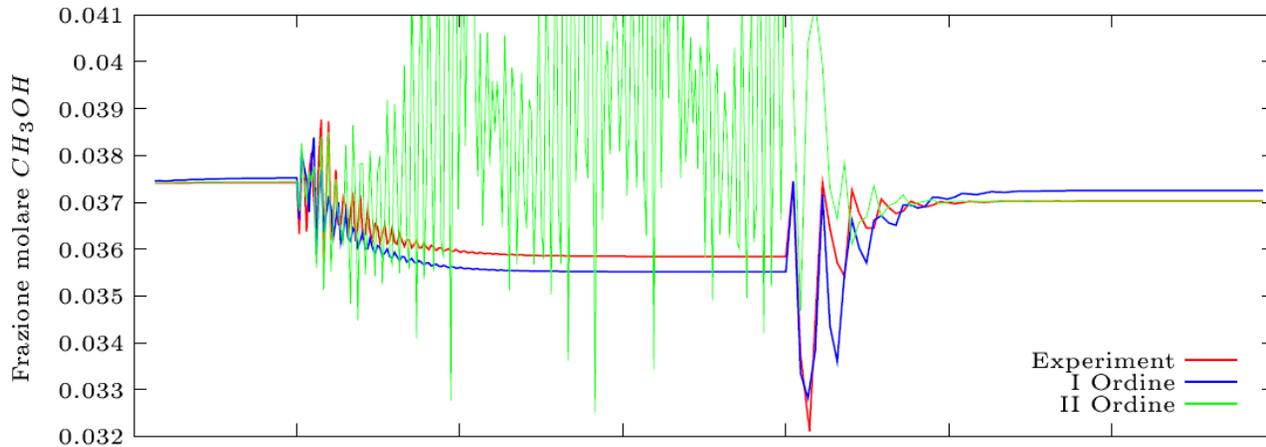
- Se i pesi ed i bias variano poco da un'iterazione all'altra;
- L'errore medio commesso;
- Il decremento dell'errore medio commesso da un'iterazione all'altra.

- Rischio esplicito di queste modalità di controllo della convergenza è l'**overfitting** della rete.



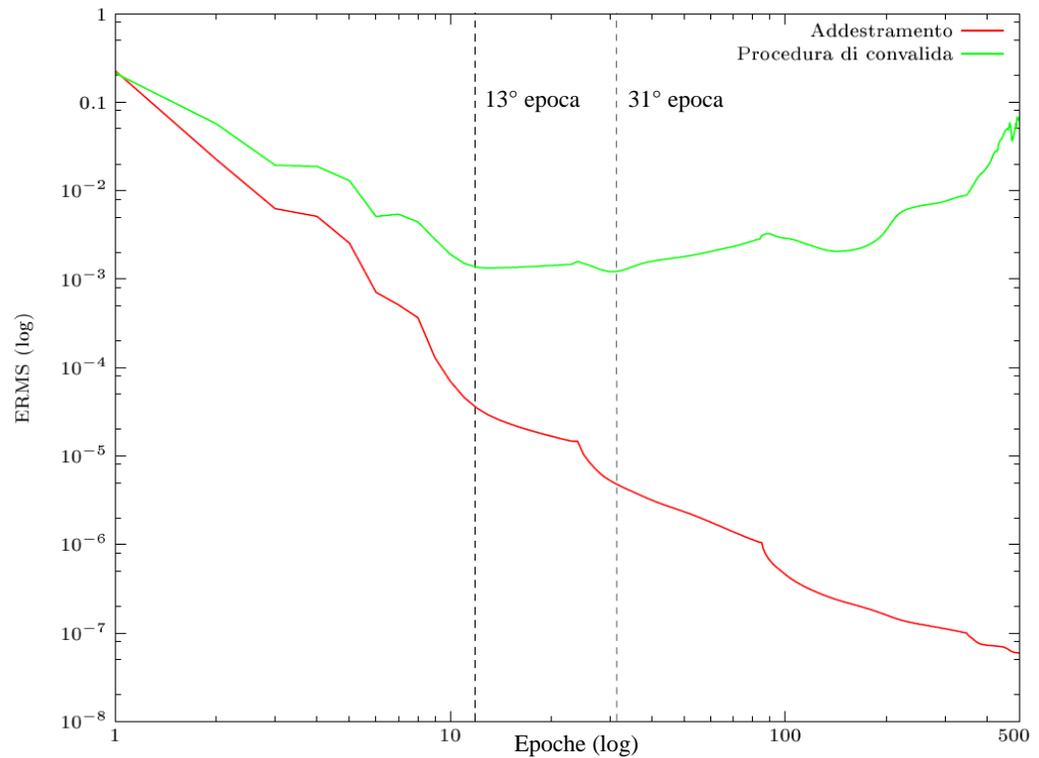
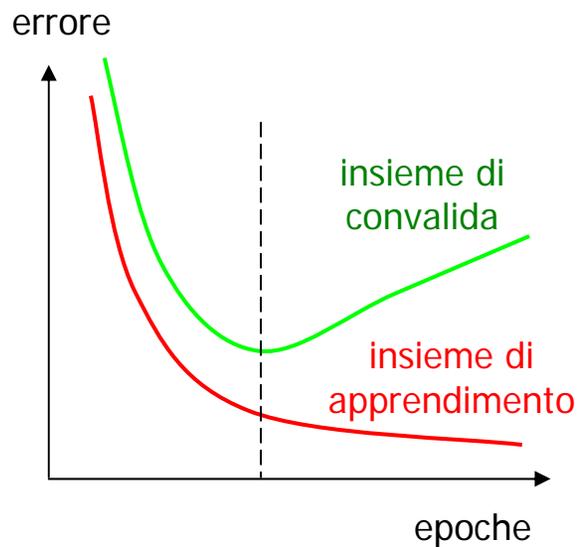
Esempio di overfitting

- Rete di reattori a letto mobile simulato per la sintesi del metanolo.



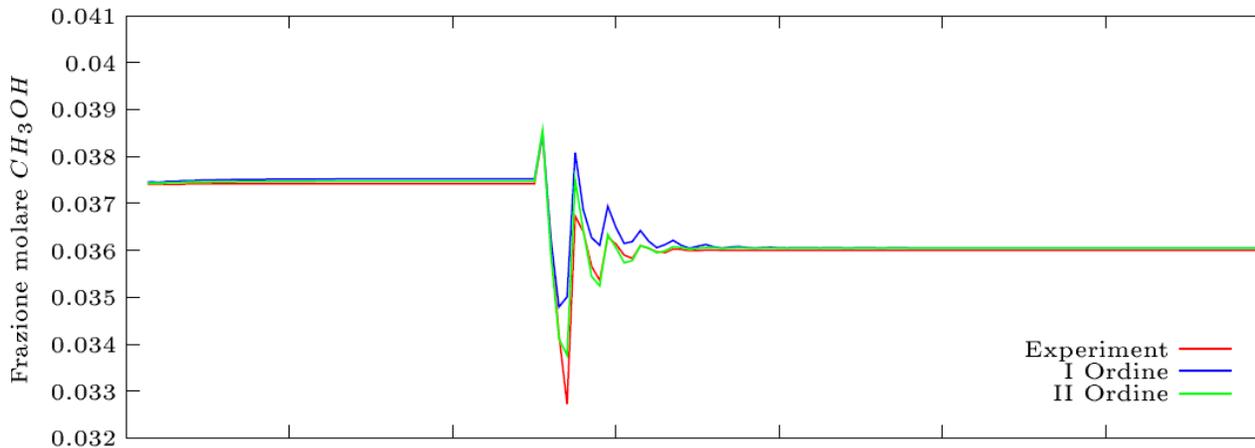
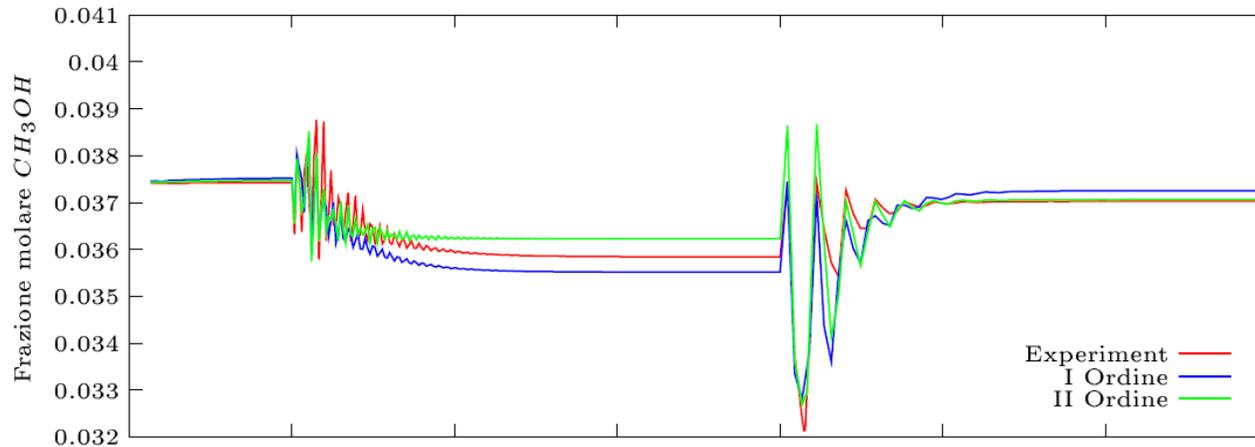
Criteri di convergenza

2. Si può focalizzare l'attenzione sul grado di identificazione relativo all'insieme dei **pattern di convalida**. Si controlla cioè l'errore medio commesso rispetto all'insieme dei dati di convalida e non di quelli di apprendimento.



Eliminazione dell'overfitting

- Rete di reattori a letto mobile simulato per la sintesi del metanolo.



Statica e dinamica

- Una rete neurale è in grado di identificare il legame tra variabili di input e variabili di output di un processo visto come scatola nera.
- Il legame input-output può fare riferimento a **condizioni stazionarie** o meglio a correlazioni funzionali che **non dipendono** né esplicitamente né implicitamente dal **tempo**.
- La rete può in alternativa identificare il **comportamento dinamico** di un processo. In questo caso dovrà essere in grado di descrivere la risposta dinamica del sistema a disturbi esterni sulle variabili di input.
- Come avviene per i sistemi ARX, ARMAX, NARX, le **variabili di input alla rete** non sono unicamente quelle di input al processo ma si hanno anche le variabili di output riferite a tempi passati.

$$\hat{\mathbf{y}}(k) = \varphi[\mathbf{y}(k-1), \mathbf{y}(k-2), \dots, \mathbf{y}(k-n_a), \mathbf{u}(k-1-\tau), \mathbf{u}(k-2-\tau), \dots, \mathbf{u}(k-n_b-\tau)]$$

Previsione del
modello

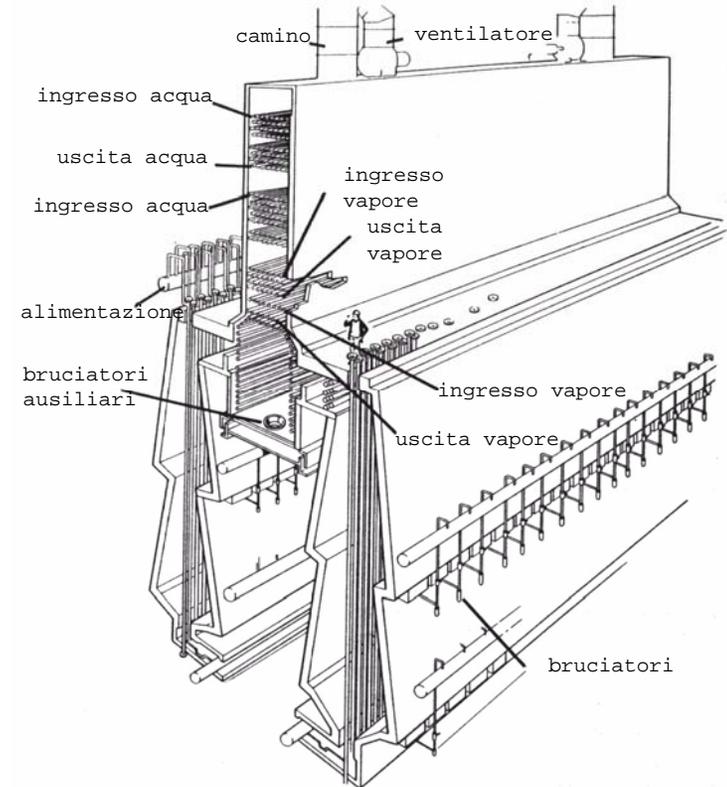
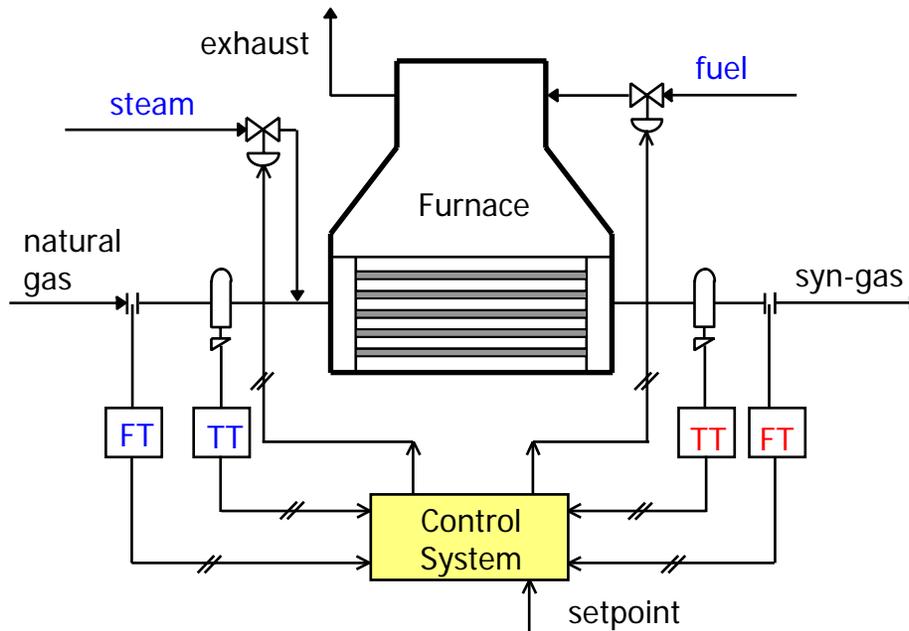
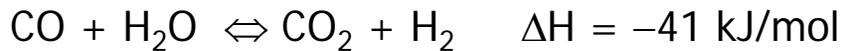
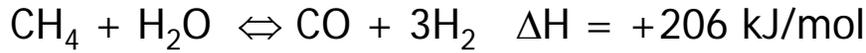
Output del
processo

Tempo di
ritardo



Identificazione dinamica di processo

- Esempio: **Processo di Steam Reforming.**



Tubi in lega speciale di Cr/Ni resistenti fino a 1150 °C.

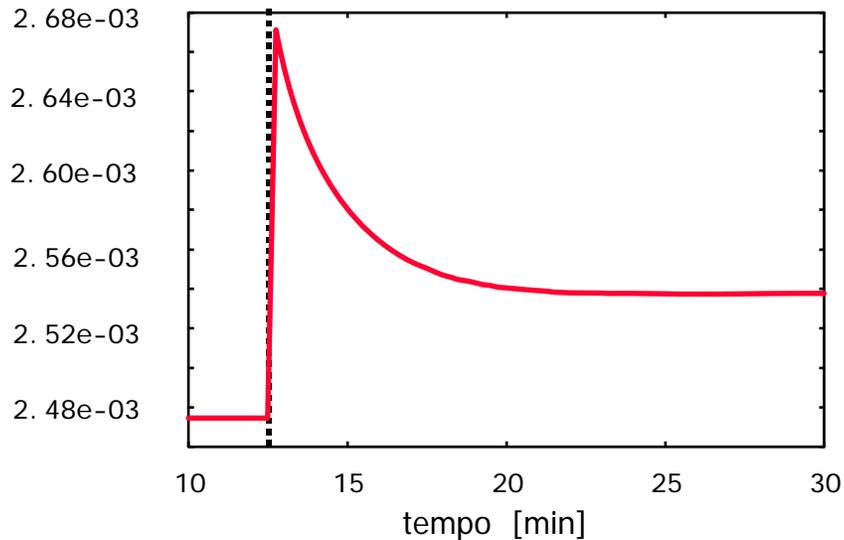
Catalizzatore in granuli: Nichel supportato su allumina

Input alla rete: variabili manipolate e disturbi misurabili

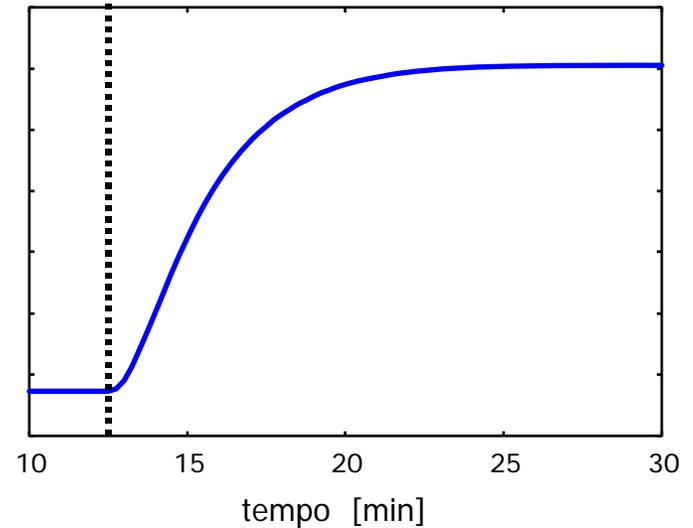
Output alla rete: variabili da controllare

Analisi del sistema ad anello aperto

Portata di idrogeno [kmol/s]



Disturbo sulla portata di secco



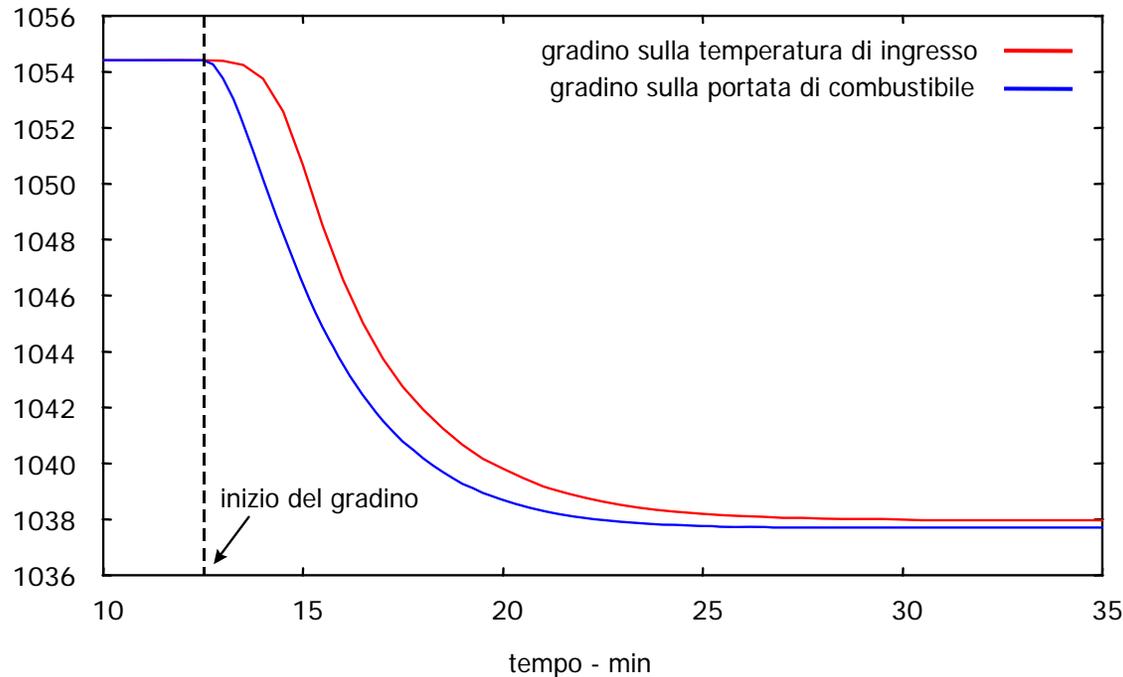
Disturbo sulla portata di fuel alla fornace

Caratteristiche del processo da identificare:

presenza di **sovrisposta**, **dinamica lenta** e **dinamica veloce**.

Analisi del sistema ad anello aperto

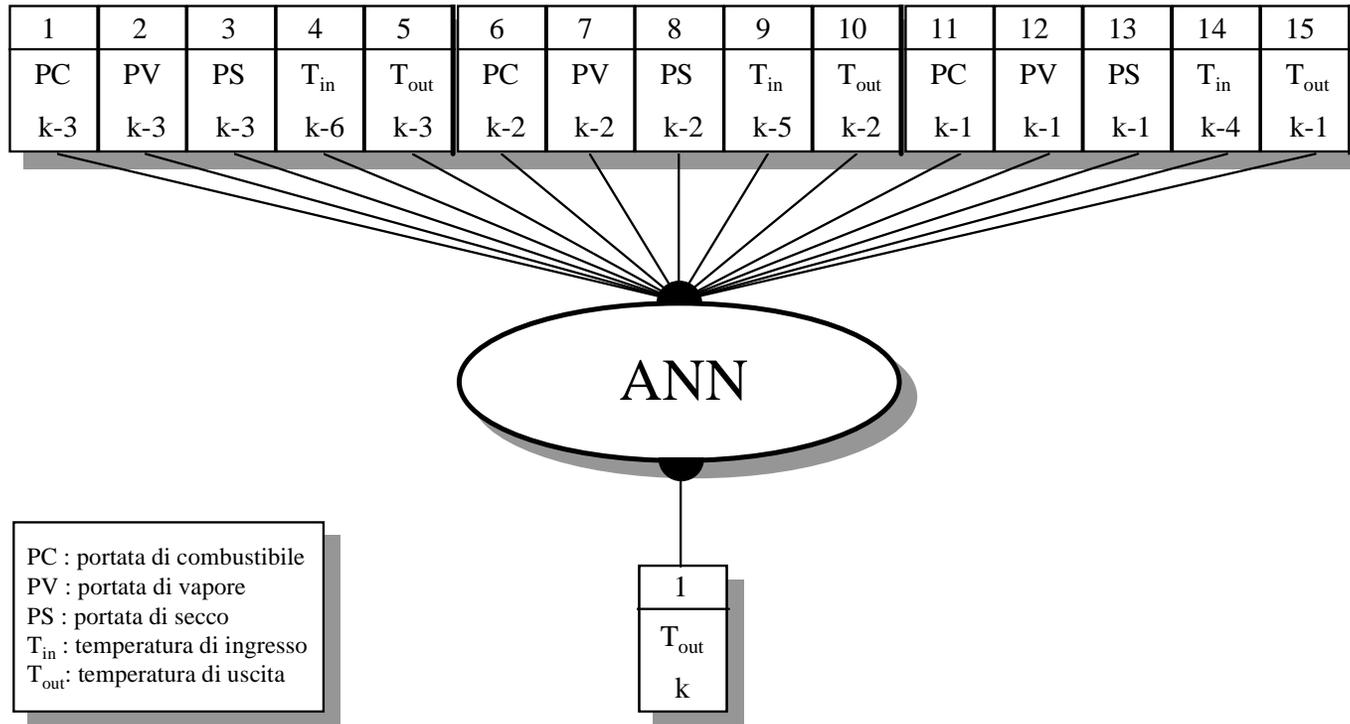
Temperatura gas di uscita [K]



N.B.: un disturbo sulla temperatura del gas in ingresso ha un effetto ritardato sulla temperatura del gas di processo in uscita dal reattore. Occorre tenerne conto nella definizione della struttura della rete neurale a livello di tempi di ritardo, τ .

Identificazione dinamica di processo

- Attributi della rete neurale: **due reti MISO**.



N.B.: la seconda rete MISO vede al posto della temperatura di uscita la variabile di output: portata di idrogeno prodotta.



Bibliografia

- Battiti R., First and Second Order Methods for Learning: Between Steepest Descent and Newton's Method, *Neural Computation*, 4, 2, 1992
- Cholewo T.J., J.M. Zurada, Exact Hessian Calculation in Feedforward FIR Neural Networks, *IEEE International Joint Conference on Neural Networks*, 1998
- Churchland P.S., T.J. Sejnowski, *The Computational Brain*, MIT Press, Cambridge, 1992
- Cybenko G., Approximation by Superpositions of a Sigmoidal Function, *Mathematics of Control Signals and Systems*, 1989
- Hagan M.T., H.B. Demuth, M. Beale, *Neural Network Design*, PWS Publishing Company, Boston, 1996
- **Haykin S., *Neural Networks: a Comprehensive Foundation*, Prentice-Hall, Upper Saddle River, New Jersey, 1999**
- Henson M.A., D.E. Seborg, *Nonlinear Process Control*, Prentice Hall, 1998
- Hilgard E.R., R.L. Atkinson, R.C. Atkinson, *Introduction to Psychology*, Harcourt Brace Jovanovich, 1979
- Hrycej T., *Neurocontrol towards an Industrial Control Methodology*, John Wiley & Sons, New York, 1997
- Narendra K.S., A.U. Levin, Identification using Feedforward Networks, *Neural Computation*, 7, 1995



Bibliografia

- Narendra K.S., K. Parthasarathy, Identification and Control of Dynamical Systems Using Neural Networks, IEEE Trans. on Neural Networks, 1, 1990
- **Principe J. C., Euliano N. R. and Lefebvre W. C., Neural and Adaptive Systems: Fundamentals through Simulations, John Wiley & Sons, New York, 2000**
- Ranga Suri N.N.R., D. Deodhare, P. Nagabhushan, Parallel Levenberg-Marquardt-based Neural Network Training on Linux Clusters - A Case Study, Indian Conference on Computer Vision Graphics and Images Processing, 2002
- Scattolini R., S. Bittanti, On the Choice of the Horizon in Long-range Predictive Control. Some Simple Criteria, Automatica, 26, 5, 1999
- **Tadé M.O., P.M. Mills, A.Y. Zomaya, Neuro-Adaptive Process Control: a Practical Approach, John Wiley & Sons, New York, 1996**
- Wan E.A., F. Beaufays, Diagrammatic Derivation of Gradient Algorithms for Neural Networks, Neural Computation, 8, 1, 1996
- Werbos P.J., The Roots of Backpropagation - From Ordered Derivatives to Neural Networks and Political Forecasting, John Wiley & Sons, New York, 1994
- Wilamowski B.M. , Neural Network Architectures and Learning, IEEE-Neural Networks Society, 2003
- Wilamowski B.M., S. Iplikci, O. Kaynak, An Algorithm for Fast Convergence in Training Neural Networks, IEEE International Joint Conference on Neural Networks, 2001

