

Reti Neurali

Cosa sono le reti neurali?

Le reti neurali sono modelli matematici composti da neuroni artificiali, ispirato alle reti neurali biologiche, che vengono utilizzate per risolvere problemi di intelligenza artificiale legati a diversi campi. Per essere più precisi le reti neurali sono modelli di calcolo matematico-informatici basati sul funzionamento delle reti neurali biologiche, ossia modelli costituiti da interconnessioni di informazioni. Una rete neurale è un sistema “adattivo” in grado di modificare la sua struttura basandosi sia su dati esterni sia su informazioni interne che si connettono e passano attraverso la rete neurale durante la fase di apprendimento.

Come ottengono informazioni?

Come le persone ottengono informazioni dall'esterno che successivamente vengono elaborate dal nostro cervello, anche le reti neurali ricevono segnali esterni su uno strato di nodi (nodi di ingresso), ognuno di questi nodi è collegato ad altri nodi interni della rete che normalmente sono organizzati a più livelli in modo che ogni singolo nodo possa elaborare i segnali ricevuti trasmettendo ai livelli successivi il risultato delle sue elaborazioni.

Le reti neurali devono avere almeno 3 strati:

- Strato di input: Tutti i dati provenienti dall'esterno arrivano lì, questo strato ha il compito di elaborarli in modo che la rete neurale possa gestirli.
- Hidden Layer: È lo strato di mezzo, più hidden layer ci sono, più la rete neurale è complessa e quindi elabora meglio le informazioni, questo livello deve proprio elaborare le informazioni che arrivano dal livello prima.
- Strato di Output: Vengono presi i risultati dell'hidden layer e vengono poi utilizzati per lo scopo che ci si era posti.

Allenamento reti neurali

Bisogna allenare le reti neurali così quando dovranno risolvere il problema per cui sono state progettate saranno in grado di farlo.

Le reti neurali imparano dall'esperienza, c'è apprendimento quando le prestazioni del programma migliorano dopo lo svolgimento di un compito o il completamento di un'azione (anche se sbagliata). Anziché scrivere il codice di programmazione attraverso il quale si dice alla macchina cosa fare, al computer vengono forniti solo dei set di dati inseriti in un generico algoritmo che sviluppa una propria logica per svolgere la funzione richiesta.

Ci sono due principali tipi di apprendimento, l'apprendimento supervisionato e quello non supervisionato.

Apprendimento Supervisionato

L'apprendimento supervisionato è una tecnica di apprendimento automatico che mira a istruire un sistema informatico in modo da consentirgli di elaborare automaticamente previsioni sui valori di uscita di un sistema rispetto ad un input sulla base di una serie di esempi ideali, costituiti da coppie di input e di output, che gli vengono inizialmente forniti.

Al computer vengono dati set di input con relativo output (corretto) con l'obiettivo che il sistema identifichi una regola generale che colleghi i dati in ingresso con quelli in uscita in modo da poter poi riutilizzare tale regola per altri compiti simili.

La complessità del modello si riferisce alla complessità della funzione che si sta tentando di apprendere. Il livello appropriato di complessità del modello è generalmente determinato dalla natura dei dati di allenamento. Se si dispone di una piccola quantità di dati o se i dati non sono distribuiti uniformemente in tutti i possibili scenari, è necessario optare per un modello a bassa complessità. Questo perché un modello ad alta complessità si adatta meglio se utilizzato su un numero ridotto di punti dati.

Apprendimento non Supervisionato

L'apprendimento non supervisionato è una tecnica di apprendimento automatico che consiste nel fornire al sistema informatico una serie di input che egli riclassificherà ed organizzerà sulla base di caratteristiche comuni per cercare di effettuare ragionamenti e previsioni sugli input successivi. Durante l'apprendimento vengono forniti all'apprendista solo esempi non annotati, in quanto le classi non sono note a priori ma devono essere apprese automaticamente.

Al sistema vengono forniti solo set di dati senza alcuna indicazione del risultato desiderato. Lo scopo di questo metodo di apprendimento è "risalire" a schemi e modelli nascosti, ossia identificare negli input una struttura logica senza che questi siano preventivamente etichettati.

Back Propagation

È un algoritmo che minimizza l'errore quadratico medio fra l'output corrente e quello desiderato.

- 1) Inizializzazione dei pesi, i pesi sono inizializzati a piccoli valori casuali.
- 2) l'input è un vettore di numeri reali x_0, x_1, \dots, x_{n-1} a cui corrisponde un vettore di output d_0, d_1, \dots, d_{m-1} . Se la rete è usata come classificatore, la classificazione è codificata usando tante unità di output quante sono le classi e ponendo l'output corrispondente alla classe corretta a 1 e tutti gli altri a 0.
- 3) Calcolo degli output correnti: dato un pattern p , procedendo a livelli si calcola l'input di ogni nodo poi si calcola l'output

$$f(z) = \frac{1}{1 + e^{-z}}$$

- 4) Modifica dei pesi:

Si procede a livelli, calcolando l'errore sul livello di output e propagandolo all'indietro fino al primo livello interno; quindi si aggiornano i pesi con la delta rule generalizzata:

che modifica il peso di una connessione in modo proporzionale all'errore compiuto dall'unità di arrivo (δ_{pi}) e alla grandezza dell'output dell'unità di partenza (o_{pj}). L'errore sulle unità viene calcolato con le formule:

$$\delta_{pi} = (d_{pi} - o_{pi})f'(\text{net}_{pi})$$

$$\delta_{pi} = \left(\sum_k \delta_{pk} w_{ik} \right) f'(\text{net}_{pi})$$

$$\text{con } f'(x) = f(x)(1 - f(x))$$

la prima calcola l'errore per il livello di output (in base alla differenza fra l'output corrente e quello desiderato), e la seconda lo propaga alle unità interne (in base all'errore delle unità del livello superiore e all'intensità delle connessioni).

Differenza reti neurali e sistemi esperti

Si tratta in entrambi i casi di varie branche dell'intelligenza artificiale, ma i sistemi esperti riproducono una persona esperta in un determinato campo, conservano e utilizzano tutta la loro conoscenza e la mettono a disposizione degli utenti.

I sistemi esperti hanno una "Base di Conoscenza", come se fosse un database di informazioni che servono al sistema per fornire una risposta (come un uomo la memorizza nel proprio cervello). I sistemi esperti hanno anche delle regole integrate, per esempio se si verifica una certa condizione allora svolgi questa azione, per esempio se viene chiesta una specifica domanda verrà data una risposta precisa oppure verrà cominciata una precisa azione.

I sistemi esperti sono dotati anche di un'interfaccia (solitamente web), questo perché possano essere utilizzati dalle persone per reperire informazioni, per esempio un utente scrive una domanda e lui gli dà la risposta.

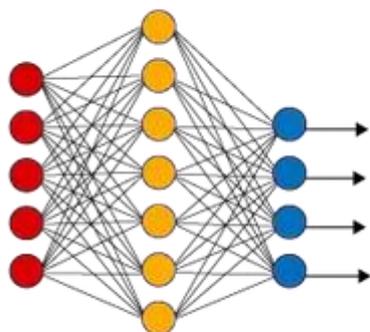
Al contrario di come molti la pensano le reti neurali e i sistemi esperti non sono la stessa cosa, ci sono cose che solo una delle due sa fare, per esempio un sistema esperto può dedurre alcuni ragionamenti, "Edoardo è un uomo, gli uomini sono mortali, Edoardo non è immortale", questo le reti neurali non possono farlo, pero possono riconoscere volti anche in immagini complesse cosa che i sistemi esperti non possono fare.

Un sistema esperto, ripercorrendo come ha applicato le regole if-then in base alla sua conoscenza riesce a spiegare come è arrivato alla soluzione, mentre le reti neurali funzionando in maniera molto più complessa non si riesce quasi mai a risalire al procedimento logico che ha portato alla soluzione. I sistemi esperti vengono programmati, mentre le reti neurali deducono le regole in modo automatico senza bisogno che vengano decise prima.

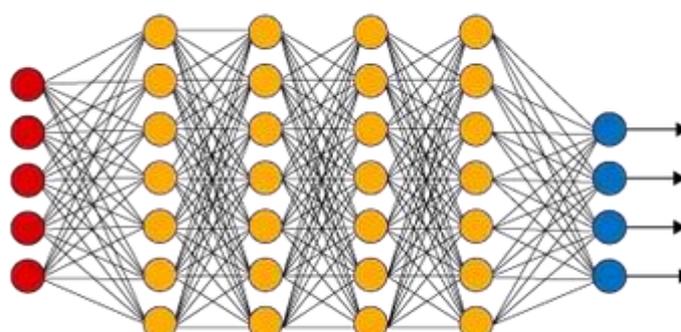
Le reti neurali potrebbero essere usate per alimentare la conoscenza dei sistemi esperti.

Architettura delle reti neurali

Simple Neural Network



Deep Learning Neural Network



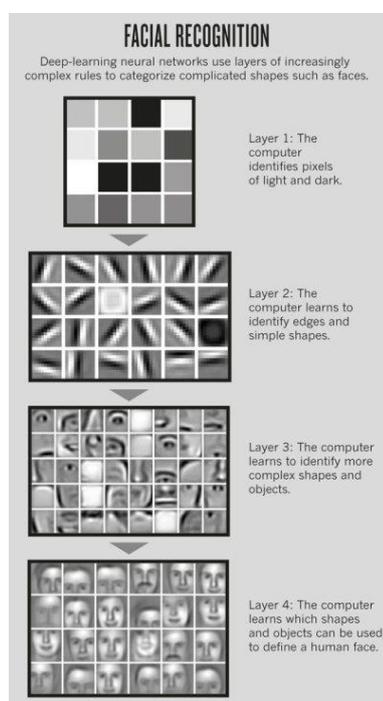
● Input Layer ● Hidden Layer ● Output Layer

Le reti neurali sono composte da minimo tre strati:

L'Hidden Layer è il livello centrale e il suo output è sempre collegato a un altro livello della rete neurale (di solito output layer).

Più la rete neurale è profonda (deep learning) più sarà in grado di svolgere il proprio lavoro in modo accurato, per esempio:

Il computer deve riconoscere i volti delle persone, essendo tutti i visi diversi uno dall'altro l'operazione che deve fare il computer è complessa perchè deve valutare migliaia di pixel di colori diversi, deve distinguere la bocca, il naso, gli occhi ecc...



In sostanza, il primo livello (dell'Hidden Layer) identifica i pixel scuri da quelli chiari, che sono assolutamente inutili se si vuole riconoscere un volto, ma sono molto utili per far riconoscere al secondo livello le forme del viso. Il terzo livello impara a riconoscere forme più complesse di semplici forme e contorni e l'ultimo livello è in grado di riconoscere un volto, se la rete neurale avesse più livelli si sarebbe svolta un'analisi più approfondita delle fotografie con probabili risultati migliori. Ogni livello trasforma sempre di più l'input di pixel in un volto.

Vantaggi reti neurali

Adatte per problemi che non chiedono risposte accurate, ma risposte approssimate con un grado di errore o di variazione.

Generalizzazione: producono buone risposte anche con input non considerati durante la creazione e l'addestramento.

Stabilità dell'output rispetto a valori di input: incompleti, con rumore, non ben noti, che accettano un grado di errore o di variazione.

Svantaggi reti neurali

Incapacità di rendere conto dell'elaborazione: non si può capire perché ha dato quel risultato specifico. Tecniche di addestramento sofisticate che richiedono molto tempo di calcolo. Serve una casistica di esempi molto ampia per ottenere un buon apprendimento e un basso errore di output.

<https://www.ai4business.it/intelligenza-artificiale/deep-learning/reti-neurali/>

https://it.wikipedia.org/wiki/Rete_neurale_artificiale

<https://vision.unipv.it/IA/aa2001-2002/10-RetiNeurali.pdf>

<https://it.quora.com/Qual-%C3%A8-la-differenza-tra-machine-learning-deep-learning-e-reti-neurali-Dove-%C3%A8-preferibile-uno-rispetto-agli-altri>

<https://www.innovationpost.it/2018/02/14/intelligenza-artificiale-deep-learning-e-machine-learning-quali-sono-le-differenze/>

<https://www.ai4business.it/intelligenza-artificiale/sistemi-esperti-cosa-sono/>

<https://vision.unipv.it/IA1/aa2005-2006/RetiNeurali.pdf>