

Introduzione alle neural networks

Federico Achini

16 Dicembre 2024



UNIVERSITÀ
DEGLI STUDI
DI MILANO

Cosa faremo oggi?

Cosa faremo oggi?

- Una (*piccola!*) introduzione teorica (diciamo un'ora).

Cosa faremo oggi?

- Una (*piccola!*) introduzione teorica (diciamo un'ora).
- Entriamo nel vivo: come riconoscere gli animali?

Cosa faremo oggi?

- Una (*piccola!*) introduzione teorica (diciamo un'ora).
- Entriamo nel vivo: come riconoscere gli animali?
- **Adesso tocca a voi!**



Cosa sono le neural networks

Algoritmo

Insieme di istruzioni che deve essere applicato per eseguire un'elaborazione o risolvere un problema (Treccani).

Cosa sono le neural networks

Algoritmo

Insieme di istruzioni che deve essere applicato per eseguire un'elaborazione o risolvere un problema (Treccani).

Praticamente è la ricetta per risolvere un problema. Ma cosa fare se non siamo capaci di scriverlo da soli?

Cosa sono le neural networks

Creiamo una macchina che costruisca gli algoritmi per noi!

Machine learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

(Tom M. Mitchell, Professor of Machine Learning Department,
School of Computer Science, Carnegie Mellon University)

Cosa sono le neural networks

Creiamo una macchina che costruisca gli algoritmi per noi!

Machine learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .

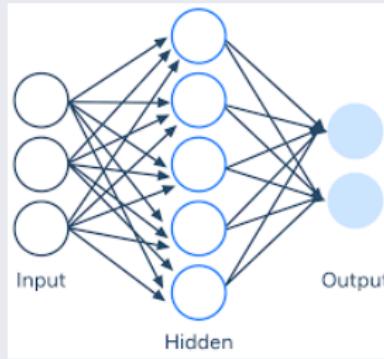
(Tom M. Mitchell, Professor of Machine Learning Department,
School of Computer Science, Carnegie Mellon University)

In pratica, E è un insieme di dati, T è il compito che vorremmo risolvere e P è l'accuratezza con cui l'algoritmo risolve il compito assegnato.

Cosa sono le neural networks

Definizione

Una **neural network** è un particolare tipo di algoritmo di machine learning, la cui struttura è ispirata a quella del cervello umano, in grado (*idealmente*) di apprendere dai dati come risolvere un determinato problema.



Supervised learning

Ci sono due grandi tipologie di apprendimento:

- **Supervised learning:** l'insieme dei dati iniziali è costituito sia da dati di input che da dati di output (**ground truth**). I dati di output vengono usati come termine di confronto durante l'apprendimento.

Si usa per:

- classificazione di oggetti in classi *note*
- previsioni di andamenti, interpolazioni, ...
- ...

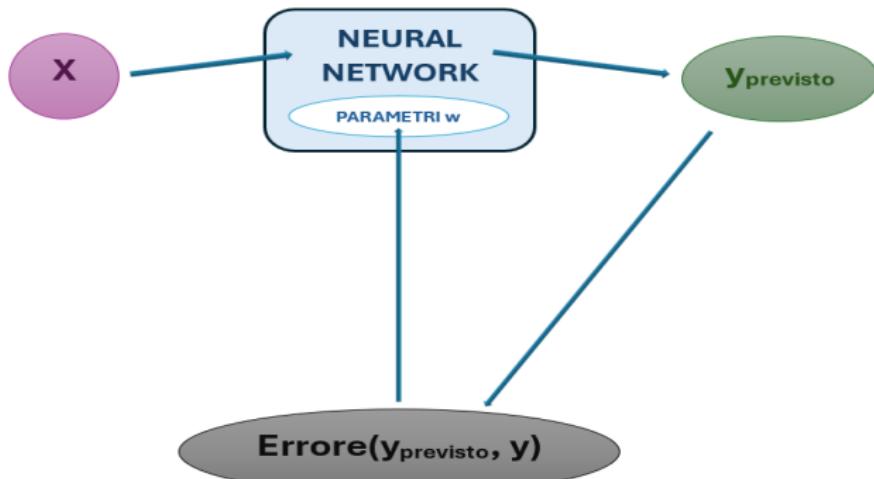
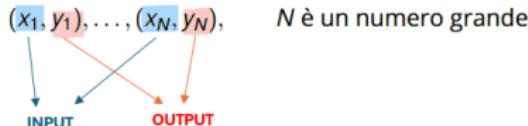
- **Unsupervised learning:** l'insieme dei dati iniziali è costituito unicamente da dati di input (non è data alcuna ground truth).

Si usa per:

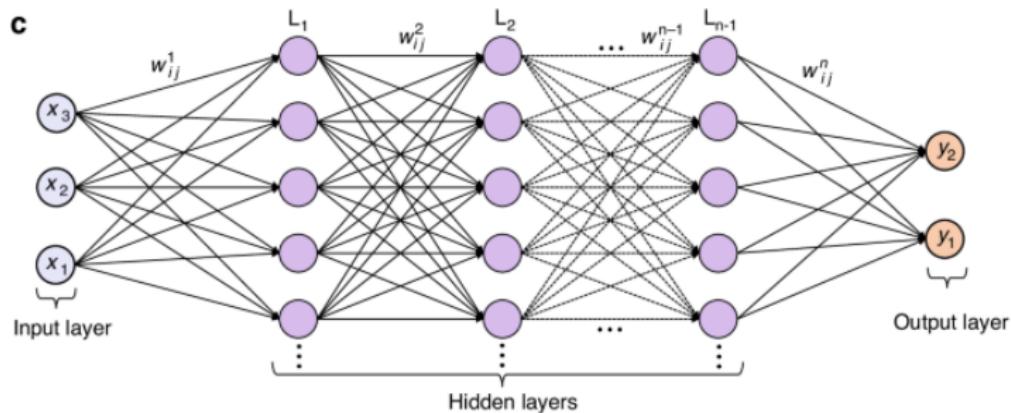
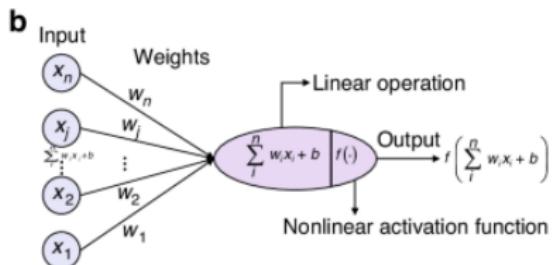
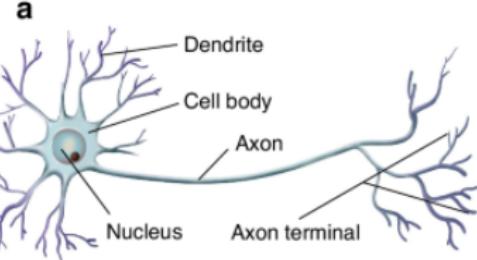
- classificazione di oggetti in classi non *note*
- ricerca di pattern all'interno dei dati.
- ...

Supervised learning

Dati iniziali



Com'è fatta una neural network?



Com'è fatta una neural network?

Si possono usare altri tipi di layers:

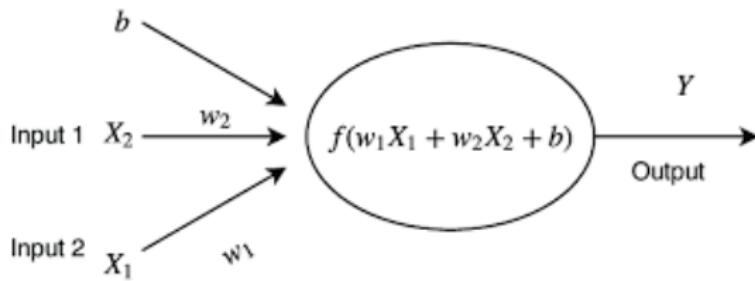
- *Densi* (quelli della slide precedente)
- *Convoluzionali*
- *Recurrent*
- ...

Quali scegliere e in che ordine posizionarli nella rete (**architettura della rete**) è una questione soprattutto di esperienza (per ora non ci sono regole, solo linee guida).

Si tratta di un campo di ricerca ancora molto attivo.

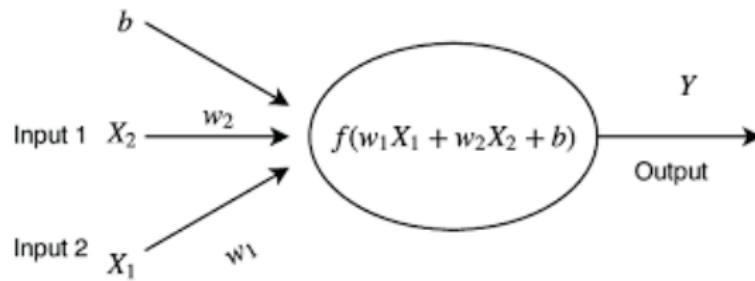
Ma come fanno ad apprendere?

Facciamo finta di avere una rete neurale costituita da un solo neurone e due soli input.



Ma come fanno ad apprendere?

Facciamo finta di avere una rete neurale costituita da un solo neurone e due soli input.



Per prima cosa ci serve un modo per calcolare l'errore. Una buona scelta è data da:

$$\text{Errore}(y_{\text{previsto}}, y) = \frac{1}{2}(y_{\text{previsto}} - y)^2.$$

Ma come fanno ad apprendere?

Osservazione

L'errore cambia in base ai parametri della neural network, ovvero l'errore è una **funzione** dei parametri w :

$$\text{Errore}(y_{\text{previsto}}, y) = \frac{1}{2}(y_{\text{previsto}}(w) - y)^2.$$

Ma come fanno ad apprendere?

Osservazione

L'errore cambia in base ai parametri della neural network, ovvero l'errore è una **funzione** dei parametri w :

$$\text{Errore}(y_{\text{previsto}}, y) = \frac{1}{2}(y_{\text{previsto}}(w) - y)^2.$$

Supponiamo di sapere che l'errore, come funzione di w , ha una derivata. Nel nostro caso la derivata rispetto a un parametro w è:

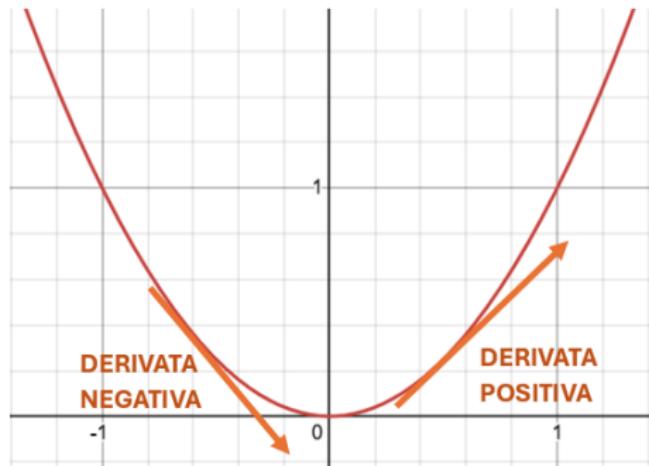
$$\begin{aligned}\frac{d\text{Errore}}{dw}(w) &= \frac{1}{2} \cdot 2 \cdot (y_{\text{previsto}}(w) - y) \cdot \frac{dy_{\text{previsto}}}{dw}(w) \\ &= (y_{\text{previsto}}(w) - y) \cdot \frac{dy_{\text{previsto}}}{dw}(w).\end{aligned}$$

Ma come fanno ad apprendere?

La derivata ci dice come cambia la funzione al cambiare di w . Per cui possiamo modificare i parametri w in maniera che l'**errore diminuisca**.

Ma come fanno ad apprendere?

La derivata ci dice come cambia la funzione al cambiare di w . Per cui possiamo modificare i parametri w in maniera che l'**errore diminuisca**.



$$w_{new} = w_{old} - \eta \cdot \frac{d\text{Errore}}{dw}(w).$$

QUESTO E' IMPORTANTE

Il parametro η è detto **learning rate**.

Ma come fanno ad apprendere?

Ho un po' barato...



Ma come fanno ad apprendere?

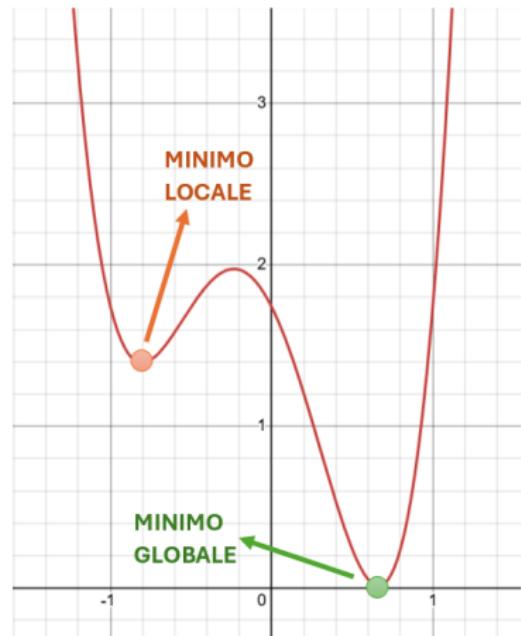
Ho un po' barato...

In generale non possiamo sapere *a priori* quale forma abbia esattamente l'errore.

Potrebbe non essere affatto una parabola!

Problema

Se ci sono altri punti di minimo, la neural network potrebbe non dare il risultato sperato.



Ma come fanno ad apprendere?

Nella pratica si usano tecniche più raffinate di quelle che avete appena visto (anche se il punto di partenza è in generale quello).

- Stochastic Gradient Descent (SGD),
- SGD with momentum,
- Root Mean Square Propagation (RMSProp),
- **Adaptive Moment Estimation (ADAM)**.

Si possono poi far variare altri fattori (**hyperparameters**), come il learning rate, il batch size, ... Tra qualche minuto, nell'esempio, vedremo alcuni dei più interessanti.

Ma come fanno ad apprendere?



Ma come fanno ad apprendere?

Ci sono altri problemi...

Underfitting

Se la rete non è sufficientemente complessa, o non viene addestrata a sufficienza, o i dati non sono molto informativi, potrebbe non essere in grado di apprendere.

Ma come fanno ad apprendere?

Ci sono altri problemi...

Overfitting

Se la rete ha molti parametri e viene addestrata troppo a lungo, potrebbe imparare a memoria i dati, rendendola poco utile!

Esami di fine anno: valutare le performances

Cosa vuol dire che una neural network ha buone performances?

Esami di fine anno: valutare le performances

Cosa vuol dire che una neural network ha buone performances?

Generalization error

Il **generalization error** è una misura dell'accuratezza con cui la neural network risolve un problema su dati che **non ha mai visto prima**.

Esami di fine anno: valutare le performances

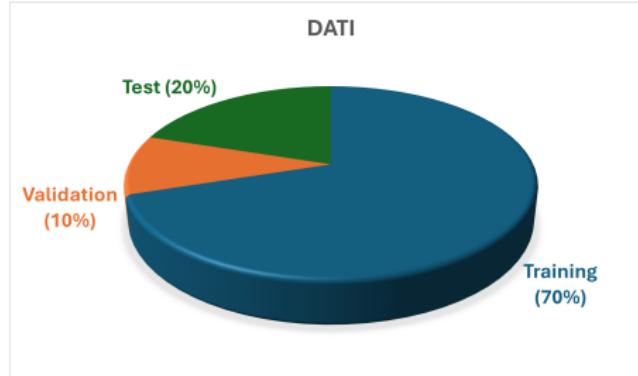
In generale si usano diverse misure a seconda del problema; per esempio:

- Accuracy (classificazione)
- Precision (classificazione)
- Recall (classificazione)
- Mean Absolute Error (MAE) (interpolazione)
- R2 score (interpolazione)
- ...

Esami di fine anno: valutare le performances

Di solito si divide l'insieme dei dati in tre sottoinsiemi:

- **Training set**: utilizzato per addestrare la rete.
- **Validation set**: utilizzato per avere un feedback sull'andamento delle performances durante l'addestramento.
- **Test set**: utilizzato per la valutazione finale delle performances.



Basta con la teoria!

Ci sarebbero molte altre cose da dire, ma... passiamo alla pratica!