**REGRESSIONE**

La **regressione è un processo statistico** che ci permette di stimare le relazioni tra variabili. Determina il**legame funzionale tra le**[variabili indipendenti e le variabili dipendenti](https://www.youmath.it/domande-a-risposte/view/6215-variabile-dipendente-variabile-indipendente.html). E’ uno dei due principali problemi dell'apprendimento supervisionato, la branca del machine learning che vuole insegnare ai computer a risolvere un determinato problema mostrandogli come questo problema è stato già risolto in passato.

La regressione corrisponde al valore medio atteso in base ai dati di input che noi immettiamo. Grazie alla regressione si può parlare di **previsione** sui valori attesi. **Intuitivamente** la regressione consente di trovare la relazione tra input e output, in modo da predire output futuri avendo a disposizione solamente l'input.

I valori di ingresso sono un vettore di dati n-dimensionale dalla quale deriva un unico valore di uscita, determinando così l'operazione di regressione.   
Ad esempio, avendo a disposizione un set di dati contenente le specifiche di diversi smartphone e il relativo prezzo di listino, potremmo utilizzare la regressione per predire a che prezzo potrà essere venduto un nuovo smartphone basandoci sulle sue specifiche.  
**Matematicamente** la regressione può essere intesa come il trovare la funzione che meglio approssima la relazione tra la variabile dipendente X (l'input) e la variabile indipendente Y (l'output). Nel caso di una regressione lineare questa funzione è un polinomio. f(x)=xw+b

**Graficamente** questa relazione lineare può essere rappresentata come una retta che passa il più vicino possibile a tutti i punti costituiti da input X e output Y.

{\displaystyle X\_{1},\ldots ,X\_{k}}{\displaystyle \mathbb {E} [Y|X\_{1},\ldots ,X\_{k}]}**I modelli di regressione**

Gli algoritmi supervisionati fanno uso di un insieme di campioni e per ogni campione abbiamo l'output corretto.

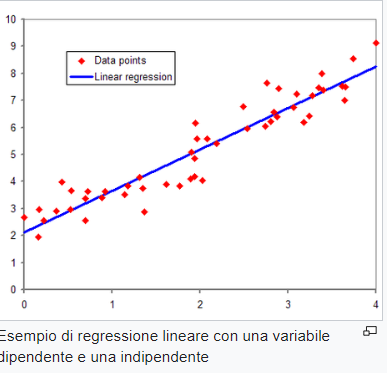
L'output corretto è associato ad ogni campione di input.

Davanti a questo tipo di situazioni abbiamo un [problema supervisionato](https://jacopokahl.com/i-tre-principali-tipi-di-machine-learning-apprendimento-supervisionato-non-supervisionato-e-per-rinforzo/).

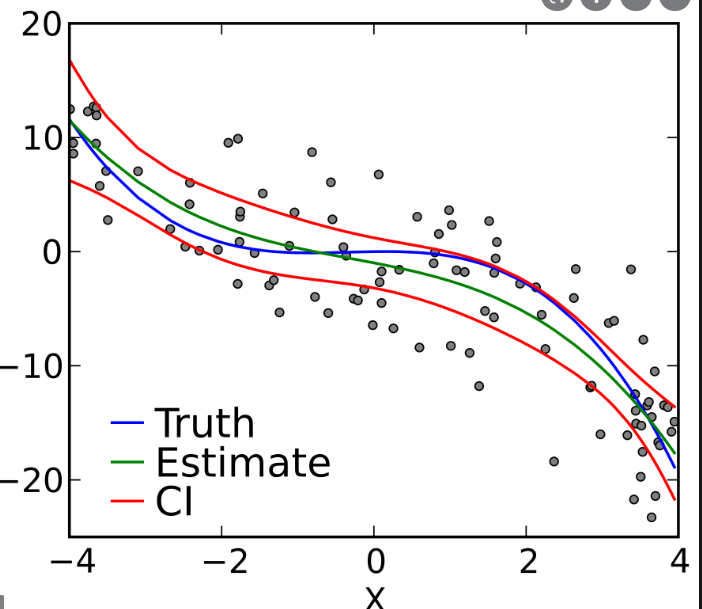
Tra i problemi supervisionati abbiamo due categorie principali: una è la regressione, l'altra è la classificazione.

La regressione consiste nello stimare un valore reale (continuo), mentre la classificazione genera un valore discreto.

1) Regressione lineare - quando la relazione tra le variabili indipendenti è di tipo lineare, abbiamo una regressione lineare.



2) Regressione polinomiale - la funzione che collega le variabili indipendenti dalle variabili dipendenti è un polinomio.

****

Dato un training set TS contenente 𝑛 pattern 𝑥1, 𝑦1 … 𝑥𝑛, 𝑦𝑛 , e la seguente dipendenza tra variabile indipendente e dipendente: 𝑦𝑖 = 𝛼 + 𝛽 ∙ 𝑥𝑖 + 𝜀𝑖 dove:  
  
𝜀𝑖 è l’errore di misura (incognito) del pattern 𝑥𝑖   
𝛼, 𝛽 sono parametri da determinar

La soluzione ai minimi quadrati (Least Square) del problema consiste nel determinare l’equazione della retta: 𝑓(𝑥) = 𝑦 = 𝛼 + 𝛽 ∙ 𝑥 che minimizza la somma dei quadrati dei residui:

Immagine che contiene testo

Descrizione generata automaticamente