

## GD DESCENT FEATURE SCALING

sabato 19 ottobre 2024 17:08

Perché il Feature Scaling è importante nel Gradient Descent?

Le caratteristiche (features) di un dataset possono avere scale molto diverse. Ad esempio, se una feature rappresenta una distanza in metri e un'altra il reddito annuale in milioni di euro, i loro valori possono essere di ordini di grandezza molto diversi. Questa disparità può causare problemi durante l'ottimizzazione con Gradient Descent.

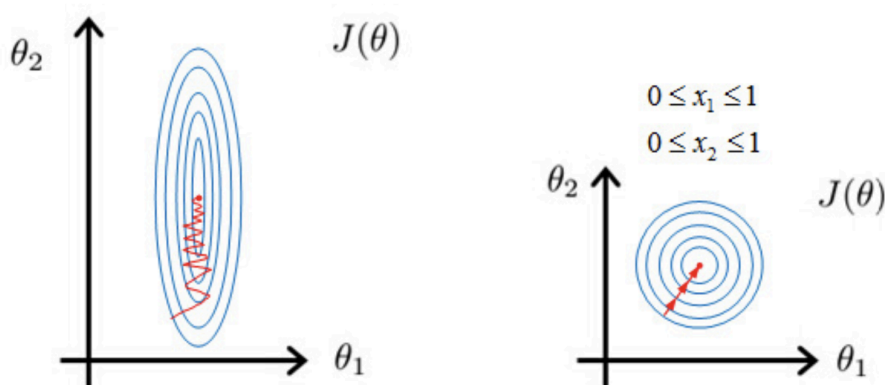
Quando le feature non sono scalate:

- Il Gradient Descent potrebbe convergere molto lentamente perché l'algoritmo deve adattarsi a gradienti molto diversi. Alcune feature potrebbero avere valori grandi e altre molto piccoli, causando "oscillazioni" lungo certe dimensioni e spostamenti minimi lungo altre.
- Alcune variabili potrebbero dominare l'aggiornamento dei parametri, rallentando il processo di ottimizzazione.

Dove sta il trucco il trucco sta a riproporzionare tali valori ES

Con il Feature Scaling, ridimensioniamo tutte le caratteristiche in un intervallo simile, per esempio tra 0 e 1, o con una media di 0 e deviazione standard 1. In questo modo:

- Il numero di stanze potrebbe passare da valori tra 1 e 10 a valori tra 0 e 1.
- L'area della casa potrebbe passare da valori tra 50 e 500 m<sup>2</sup> a valori tra 0 e 1.



PARLIAMO ORA DELLA MIN-MAX NORMALIZATION

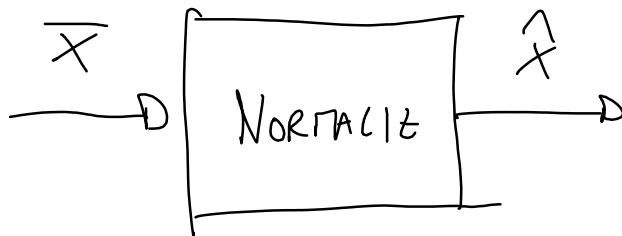
QUINDI QUANDO SI PARLAMO DI MIN-MAX SI PARLA DI NORMALIZZAZIONE.

MIN-MAX = NORMALIZZAZIONE  $[e, b]$

Def:

$$X_{\text{SCALED}} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} (b - e) + e$$

Esempio:  $\bar{X} = [1, 2, 3]$   $\rightarrow$   $X_{\text{scaled}}$



$$\hat{X} = [0, 0.5, 1]$$

$$\text{NORMALIZZAZIONE} = \frac{(x - 1)}{(3 - 1)} \cdot (1 - 0) + 0$$

$$\text{Norm} = \frac{x - 1}{2}$$

ottenendo  $\hat{X}$

Problema degli Outlier

Uno dei problemi principali di questo approccio è che gli outlier possono avere un effetto negativo:

- Se il dataset contiene valori estremamente grandi o piccoli (outlier), l'intervallo di normalizzazione sarà molto ampio.
- Questo potrebbe ridimensionare i valori più comuni in una scala molto piccola, rendendo difficile per l'algoritmo di apprendimento

sfruttarli efficacemente.



## Problemi con Nuovi Valori

In fase di predizione, potresti incontrare nuovi input che sono più grandi del massimo o più piccoli del minimo che hai usato durante la normalizzazione. In tal caso, i nuovi valori cadranno fuori dall'intervallo  $[a, b]$ .



## Soluzioni Alternative

Per evitare questi problemi, potresti considerare tecniche come lo Standard Scaling (che usa media e deviazione standard per riscalare i dati), che è meno sensibile agli outlier.

## GD Z-SCORE NORMALIZATION

Quando si utilizza la discesa del gradiente per ottimizzare modelli di apprendimento automatico, è fondamentale applicare il feature scaling per garantire che tutte le caratteristiche abbiano lo stesso peso nel processo di ottimizzazione. Ecco perché il Z-score normalization è spesso raccomandato:

### Z-score Normalization e Discesa del Gradiente

#### 1. Definizione di Z-score Normalization:

- La normalizzazione Z-score, o standardizzazione, trasforma i dati in modo che abbiano una media di 0 e una deviazione standard di 1.

1. Viene calcolata con la formula:

$$Z = \frac{x - \mu}{\sigma}$$

Dove:

- $x$  è il valore originale,
- $\mu$  è la media del dataset,
- $\sigma$  è la deviazione standard del dataset.

**Uniformità delle Scale:** La discesa del gradiente converge più rapidamente quando tutte le caratteristiche sono sulla stessa scala. Le funzioni di costo sono più lisce e ciò riduce il rischio di oscillazioni durante l'ottimizzazione.

- **Meno Influenza degli Outliers:** Sebbene gli outliers possano comunque influenzare la media e la deviazione standard, Z-score

normalization tende a mitigare il loro impatto rispetto ad altre tecniche di scaling, come la normalizzazione Min-Max. Questo è utile perché gli outliers possono deviare i gradienti e rallentare il processo di convergenza.

- Nessun Bisogno di Min e Max: A differenza della normalizzazione Min-Max, che richiede di conoscere i valori minimi e massimi delle caratteristiche, Z-score normalization non ha bisogno di queste informazioni. Questo è particolarmente vantaggioso in situazioni in cui i dati non sono stabili o contengono outliers.