GRADIENT DISCENDETE

venerdì 11 ottobre 2024 18:13

O ANCHE CHIAMATO STEEPEST DESCENT

il gradiente discendente (in inglese “Gradient Descent”) è un algoritmo di ottimizzazione molto utilizzato nell’ambito del machine learning e dell’ottimizzazione matematica. Il suo scopo è trovare il minimo di una f. unzione di costo (o funzione obiettivo) regolando iterativamente i parametri del modello in modo da ridurre il valore di questa funzione.

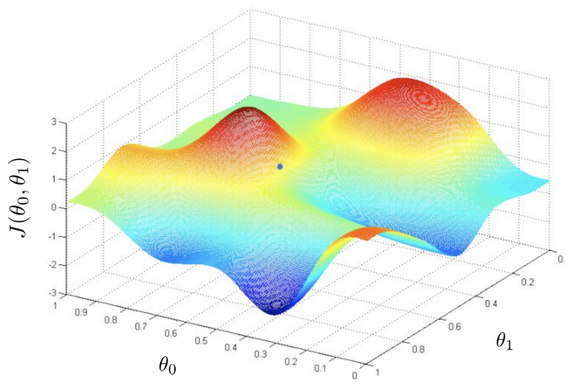
La nostra funzione di costo la tratteremo come una funzione generica

$$J(\theta_0, \theta_1) \xrightarrow{n} J(\theta_0, \theta_1 \dots \theta_n)$$

IL NOSTRO OBIETTIVO E' APPUNTO MINIMIZZARE TALE FUNZIONE

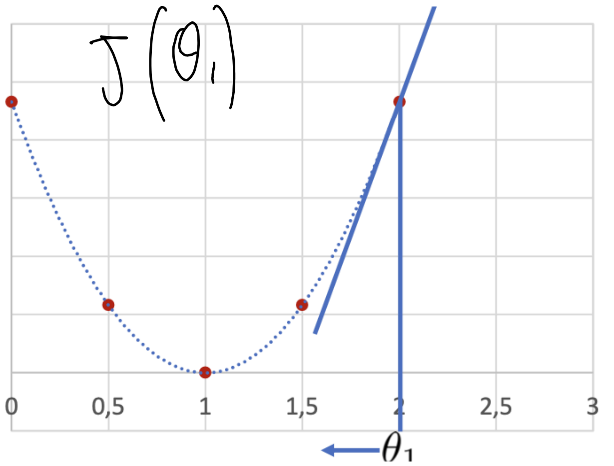
$$\min_{\theta_0,\theta_1} \left(J(\theta_0, \theta_1) \right) \qquad \min_{\theta_0,\theta_1,\theta_n} \left(J(\theta_0, \theta_1 \dots \theta_n) \right)$$

Inizializzare i valori θ_0 e θ_1 Per ridurre $J(\theta_0, \theta_1)$
E li cambieremo fino a raggiungere il minimo



Ripetere tale algoritmo

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$



$$\theta_1 = \theta_1 - \alpha \underbrace{\frac{\partial}{\partial \theta_1} J(\theta_1)}_{\geq 0}$$

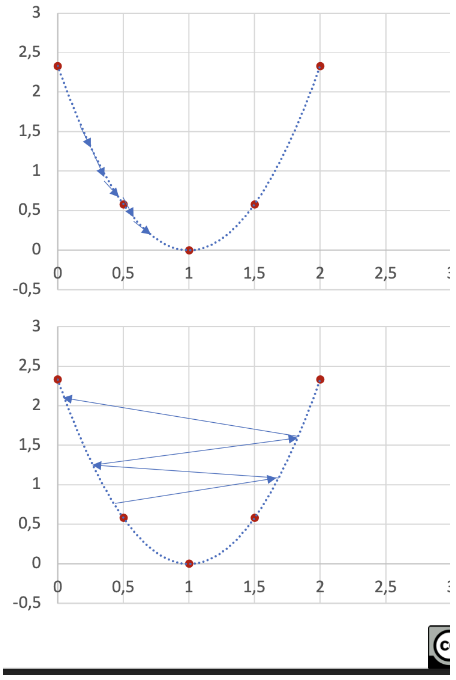
Nella parte superiore dell'immagine, vediamo la derivata parziale maggiore di 0. Questo significa che il valore del parametro è situato sul lato destro del minimo, dove la funzione di costo sta aumentando.

α = learning rate DETERMINA DI QUANTO SI AGGIORNA θ

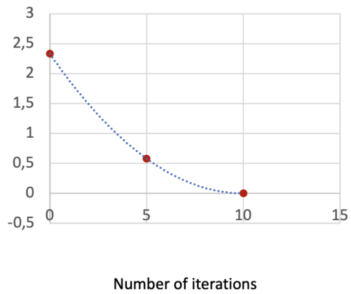
DI SOLITO ALFA VIENE SCELTO TRA 10^{-4} e 1

SE ALFA E' TROPPO PICCOLO tale metodo diventa molto lento

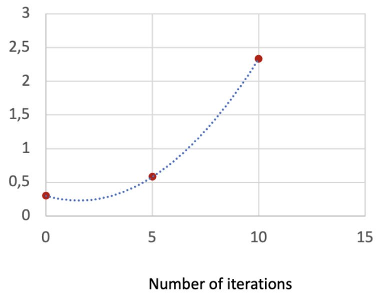
Se alfa è troppo grande, il metodo può divergere



Un altro avvenimento che può accadere è il seguente

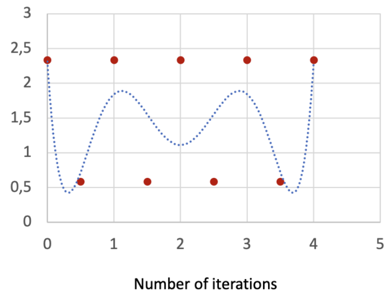


Gradient descent is working



Gradient descent is NOT working

Use smaller alpha



Gradient descent is NOT working

use smaller alpha

GRADIENT DISCENDET FOR LINEAR REGRESSION

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Qui, h_θ è l'ipotesi o la predizione del modello per un dato valore di input x. θ_0 è l'intercetta e θ_1 è il coefficiente angolare della retta di regressione.

INTERCETTA: il punto in cui la retta di regressione interseca l'asse y .

La funzione di costo che voglio minimizzare è la seguente

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)^2$$

m = esempi totali

$h_\theta\left(x^{(i)}\right)$ = E' la previsione del modello per il modello $x^{(i)}$

$y^{(i)}$ = È il valore reale al punto $x^{(i)}$

$h_\theta\left(x^{(i)}\right) - y^{(i)}$ = MSE tra il vero e la previsione

N.B θ_1 e θ_2 all'inizio si prendono per valori cas

$$\boxed{\frac{\partial}{\partial \theta_j} J(\theta_1, \theta_2) > 0}$$

VUOL DIRE CHE DEVE ESSERE RIDOTTA E QUINDI E' TROPPO BIG

$$\boxed{\frac{\partial}{\partial \theta_j} J(\theta_1, \theta_2) < 0}$$

VUOLE DIRE CHE θ_j E' TROPPO PICCOLO E QUINDI AUMENTATA

DIMOSTRAZIONE

$$\frac{\partial}{\partial \theta_j} J(\theta_1, \theta_2) = \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$= \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(\underbrace{\theta_0 + \theta_1 x^{(i)}}_{} - y^{(i)}\right)^2$$

$h_\theta(x) = \theta_0 + \theta_1 x^{(i)}$ ABBIAMO SOSTITUITO PER DEF

ORA PREDIAMO IL TERMINE DOVE DOBBIAMO FARE LA DERIVATA PARZ

$j = 0$ cosa succede?

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_0} \underbrace{\left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2}_{}$$

$$= \frac{1}{2m} \sum_{i=1}^{m} 2 \cdot \left(h_\theta\left(x^{i}\right) - y^{(i)}\right)$$

RICORDIAMO CHE NOI DERIVIAMO PER θ_0 OTT:

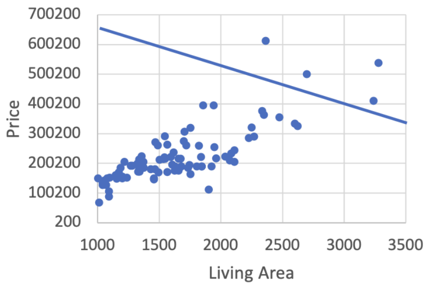
$$\frac{1}{m} \cdot \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \qquad \boxed{j:0}$$

per $\boxed{j=1}$ $= \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x^{(i)}$

Perché derivo per θ_1

$h_\theta(x)$

(function of x, with θ_0, θ_1 fixed)



$J(\theta_0, \theta_1)$

(function of θ_0, θ_1, an aggregate function over all x's)