NORMAL EQUATIONS

sabato 19 ottobre 2024 17:29

La "Normal Equation" è una formula utilizzata per risolvere i problemi di regressione lineare nel caso in cui si voglia minimizzare l'errore quadratico medi tra le predizioni e i valori osservati. Vediamo cosa rappresentano le varie componenti nell'immagine:

$$X = \begin{bmatrix} x^{(1)T} \\ x^{(2)T} \\ x^{(3)T} \\ x^{(m)T} \end{bmatrix} \qquad y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(i)} \\ y^{(m)} \end{bmatrix}$$

PER IL MOMENTO PRENDIAMO IN CONSIDERAZIONE LA FUNZIONE DI COSTO :

Vettore RIGA , Vettore Colonne

$$J(\theta) = \frac{1}{2}(X\theta - y)^T \cdot (X\theta - y)$$

ORA DOBBIAMO MINIMIZZARE TALE FUNZIONE DI COSTO E TALE MINIMIZZAZIONE LA SI FA CON IL GRADIENTE

FACENDO LA DERIVATA OTTENIAMO

Norma

$$\frac{\partial}{\partial\theta}\left[\frac{1}{2}(X\theta - y)^T \cdot (X\theta - y)\right]$$

N.B $\|v\|^2 = v^T v$

POICHE' SI TRATTA DI UN PRODOTTO SCALARE OTTENIAMO

$$\frac{\partial J(\theta)}{\partial\theta} = (X\theta - y)^T \left(\frac{\partial(X\theta - y)}{\partial\theta}\right)$$

$$= X^T(X\theta - y)$$

IL FATTORE	$\frac{1}{2}$	AIUTA IN QUANTO AVREMO :

$$2X^T(X\theta - y)$$

si va ad elidere

IN DEFINITIVA AVREMO:

$$\nabla J(\theta) = X^T X\theta - X^T y = 0$$

ORA SE VOGLIAMO OTTENERE θ

$$X^T X\theta - X^T y = 0$$

$$= X^T X\theta = X^T y$$

$$= \hat{\theta} = \frac{X^T y}{X^T X} = (X^T X)^{-1} \cdot X^T y = \frac{1}{(X^T X)} \cdot X^T y$$

DAL PUNTO DI VISTA COMPUTAZIONALE SI HA $O(m^3)$ EQUIVALE A 3 CICLI FOR UNO DENTRO L'ALTRO.

$$\mathbf{X}\theta - \mathbf{y} = \begin{bmatrix} 1 & x_1^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & \dots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(m)} & & x_n^{(m)} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} - \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} = \begin{bmatrix} h_\theta(\mathbf{x^{(1)}}) - y^{(1)} \\ h_\theta(\mathbf{x^{(2)}}) - y^{(2)} \\ \vdots \\ h_\theta(\mathbf{x^{(m)}}) - y^{(m)} \end{bmatrix}$$

NOI DOBBIAMO TROVARE IL MINIMO, CIOE IL PUNTO DOVE IL GRADIENTE HA VALORE NULLO

$$\nabla J(\theta) = 0$$

$$\nabla J(\theta) = \begin{bmatrix} \sum_{i=1}^{m}(h_\theta(\mathbf{x^{(i)}}) - y^{(i)}) \\ \sum_{i=1}^{m}(h_\theta(\mathbf{x^{(i)}}) - y^{(i)})x_1^{(i)} \\ \vdots \\ \sum_{i=1}^{m}(h_\theta(\mathbf{x^{(i)}}) - y^{(i)})x_n^{(i)} \end{bmatrix}$$

$$\begin{aligned} \nabla J(\theta) &= \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(m)} \end{bmatrix} \begin{bmatrix} h_\theta(\mathbf{x} \\ h_\theta(\mathbf{x} \\ \\ h_\theta(\mathbf{x} \end{bmatrix} \\ &= \mathbf{X}^T(\mathbf{X}\theta - \mathbf{y}) \\ &= \mathbf{X}^T\mathbf{X}\theta - \mathbf{X}^T\mathbf{y} \end{aligned}$$

$$\begin{aligned} \nabla J(\theta) &= \mathbf{X}^T\mathbf{X}\theta - \mathbf{X}^T\mathbf{y} = 0 \\ \hat{\theta} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \end{aligned}$$