2. GRADIEN DESCENT LOGISTIC REGRESSION

giovedì 24 ottobre 2024 15:55

Obiettivo: Minimizzare la funzione di costo J(θ), ovvero trovare i valori ottimali dei parametri
θ che minimizzano J(θ)

$$\hat{\theta} = \arg\min_{\theta} J(\theta)$$

Aggiornamento dei pesi: I parametri vengono aggiornati secondo la regola:

$$\theta_K = \theta_K - \alpha \frac{\partial J(\theta)}{\partial \theta_K}$$

TORNANDO ALLA NOSTRA FUNZIONE LOGISTICA

$$g(z) = \frac{1}{1 + e^{-z}}$$

$$g'(z) = \frac{\partial}{\partial z}\left(\frac{1}{1 + e^{-z}}\right)$$

RICORDANDO CHE $\frac{d}{dx}\left[\frac{f}{g}\right] = \frac{f'g - f \cdot g'}{[f(x)]^2}$

ELIMINA

$$g'(z) = \frac{\underline{0 \cdot (1 + e^{-z})} - (-e^{-z} \cdot 1)}{[1 + e^{-z}]^2}$$

$$g'(z) = \frac{e^{-z}}{(1 + e^{-z})(1 + e^{-z})} = \frac{e^{-z}}{1 + e^{-z} + e^{-z} + e^{-2z}} \checkmark$$

COMUNQUE SI VA A CALCOLARE g(z)(1-g(z)=g'(z)

$$\frac{1}{1 + e^{-z}} \cdot \left(1 - \frac{1}{1 + e^{-z}}\right) = \frac{e^{-z}}{(1 + e^{-z})^2}$$

$$g \cdot (1 - g) = g'$$ PONIAMO [illegible]

e^{-z}

$$\frac{1}{1+e^{-z}}\left(1-\frac{1}{1+e^{-z}}\right)=\frac{\quad}{(1+e^{-z})^2}$$

$$\frac{1}{1+e^{-z}}-\left(\frac{1}{(1+e^{-z})^2}\right)=\frac{e^{-z}}{(1+e^{-z})^2}$$

✓

$$\frac{1}{1+e^{-z}}-\frac{1}{(1+e^{-z})^2}=\frac{e^{-z}}{(1+e^{-z})^2}$$

$$(1+e^{-z})^2\cdot\frac{(1+e^{-z})-1}{(1+e^{-z})^2}=\frac{e^{-z}}{(1+e^{-z})^2}\cdot(1+e^{-z})^2$$

$$\hookrightarrow (1+e^{-z}-1)=e^{-z}$$

$$1+e^{-z}=e^{-z}+1$$ ✓

WEIGHT UPDATE DERIVATION | DERIVAZIONE DEI PESI DELL'AGGIORNAMENTO DI REGRESSIONE LOGISTICA

$$g'(z)=g(z)(1-g(z))$$

NOI SAPPIAMO che $h_\theta(x)=g(\theta^T x)$

Ora dobbiamo porre in essere alcune relazioni.

La moltiplicazione che vedi nella derivata di $h_\theta(x)$ rispetto a θ_k avviene perché stiamo applicando la **regola della catena** per calcolare la derivata di una funzione composta.

In dettaglio:

- $h_\theta(x)$ è la funzione sigmoide applicata a $\theta^T x$, ossia:

$$h_\theta(x)=g(\theta^T x)$$

  dove $g(z)$ è la funzione sigmoide e $z=\theta^T x$. Per calcolare la derivata rispetto a θ_k, dobbiamo considerare il fatto che $h_\theta(x)$ dipende da θ_k attraverso $\theta^T x$.

- La **regola della catena** ci dice che, quando abbiamo una funzione composta come $h_\theta(x)=g(\theta^T x)$, la derivata rispetto a θ_k si calcola come:

$$\frac{\partial h_\theta(x)}{\partial \theta_k}=\frac{\partial g(\theta^T x)}{\partial z}\cdot\frac{\partial z}{\partial \theta_k}$$

  dove $z=\theta^T x$.

$$\frac{\partial h_\theta(x)}{\partial \theta_k} = h_\theta(x)(1-h_\theta(x)) \cdot \frac{\partial(\theta^T x)}{\partial \theta_k}$$

Ma da relazioni precedenti abbiamo che :

$$\frac{\partial(\theta^T x)}{\partial \theta_k} = x_k$$

Ricordando inoltre che $\theta^T x = \theta_1 x_1 + \theta_2 x_2$

AVREMO $\frac{\partial \theta^T x}{\theta_k} = x_k$

OTTENENDO COSÌ

$$\frac{\partial h_\theta(x)}{\partial \theta_k} = h_\theta(x) \cdot (1-h_\theta(x))\, x_k$$

La moltiplicazione è quindi il risultato della combinazione della derivata della sigmoide e della derivata della funzione lineare $\theta^T x$

INTEPRETAZIONE PROBABILISTICA DELLA FUNZIONE DI COSTO INERENTE FUNZIONE DI ERRORE DI ENTROPIA INCROCIATA.

$$\frac{\partial J(\theta)}{\partial \theta_k} = -\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} \underbrace{\frac{\partial}{\partial \theta_k}(\log h_\theta(\mathbf{x}^{(\mathbf{i})}))}_{1} + (1-y^{(i)}) \underbrace{\frac{\partial}{\partial \theta_k}(\log(1-h_\theta(\mathbf{x}^{(\mathbf{i})})))}_{2} \right)$$

La funzione di costo per la regressione logistica è definita come la media della log-verosimiglianza negativa. Quando si cerca di minimizzarla usando la discesa del gradiente, dobbiamo calcolare la derivata rispetto ai parametri θ_k

Ora possiamo trattare le due derivate separatamente

$$\frac{\partial}{\partial \theta_k} \log h_\theta(x^{(i)}) = \frac{1}{h_\theta(x^{(i)})} \cdot \frac{\partial h_\theta(x^{(i)})}{\partial \theta_k}$$

Ma noi sappiamo che:

$$\frac{\partial h_\theta(x^{(i)})}{\partial \theta_k} = \underbrace{h_\theta(x^{(i)})\left(1-h_\theta(x^{(i)})\right)x_k^{(i)}}$$

Sostituendo otteniamo:

$$\frac{\partial}{\partial \theta_k} \log h_\theta(x^{(i)}) = \left(1-h_\theta(x^{(i)})\right)x_k^{(i)} \qquad 1$$

2

$$\frac{\partial}{\partial \theta_k}\left(\log\left(1-h_\theta(x^{(i)})\right)\right) = ?$$

Bisogna fare anche la derivata all'interno del log otteniamo:

$$= \frac{1}{1-h_\theta(x^{(i)})} \cdot \frac{\partial}{\partial \theta_k}\left(1-h_\theta(x^{(i)})\right)$$

Ma dalla relazione predente otteniamo:

$$\frac{\partial \log(1-h_\theta)}{\partial \theta_k} = \frac{1}{1-h_\theta(x^{(i)})} \cdot \left(-h_\theta(x^{(i)})\right)\left(1-h_\theta(x^{(i)})\right)x_k^{(i)} =$$

$$= -h_\theta(x^{(i)})\,x_k^{(i)}$$

IN DEFINITIVA OTTENIAMO:

$$\frac{\partial J(\theta)}{\partial \theta_k} = -\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)}\left(1-h_\theta(x^{(i)})\right)x_k^{(i)} + \left(1-y^{(i)}\right)\left(-h_\theta(x^{(i)})\right)x_k^{(i)}\right)$$

FATTORIZZANDO

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

ANDANDO A MOLTIPLICARE TERMINE A TERMINE

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