Minimizzazione della funzione di errore nella regressione lineare

Basso Kevin De Fina Giuseppe Mantoan Matteo
Rampazzo Filippo Vigolo Davide

March 24, 2025

Esercizio 15 (*) foglio 1 (1-6 Marzo 2025) Corso Probabilità e statistica a.a. 2024-2025

Data la funzione di errore:

$$\varphi(a, b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2 ,$$

vogliamo trovare (a^*, b^*) che la minimizzano.

1. Calcolo delle derivate parziali

Le derivate parziali rispetto ad a e b sono:

$$\frac{\partial \varphi}{\partial a} = -2 \sum_{i=1}^{n} x_i (y_i - (ax_i + b)) = 0,$$

$$\frac{\partial \varphi}{\partial b} = -2 \sum_{i=1}^{n} (y_i - (ax_i + b)) = 0.$$

2. Sistema di equazioni

Dividendo per -2 e riorganizzando:

$$\begin{cases} \displaystyle\sum_{i=1}^{n} x_i (y_i - (ax_i + b)) = 0, & \text{(Equazione 1)} \\ \displaystyle\sum_{i=1}^{n} (y_i - (ax_i + b)) = 0. & \text{(Equazione 2)} \end{cases}$$

3. Soluzione per b (Equazione 2)

Dall'Equazione 2:

$$\sum_{i=1}^{n} (y_i - (ax_i + b)) = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} b = 0$$

$$\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i - nb = 0$$

$$b = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{a}{n}\sum_{i=1}^{n} x_i.$$

$$b = \overline{y} - a\overline{x}.$$

4. Soluzione per a (Equazione 1)

$$\sum_{i=1}^{n} (y_i - (ax_i + b)) = 0$$

Sostituendo $b = \bar{y} - a\bar{x}$ nell'Equazione 1:

$$\sum_{i=1}^{n} x_i \left(y_i - ax_i - (\bar{y} - a\bar{x})\right) = 0.$$

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) - a\sum_{i=1}^{n} x_i(x_i - \bar{x}) = 0.$$

Riscriviamo le somme come:

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + \bar{x}\sum_{i=1}^{n}(y_i - \bar{y}),$$

dove $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$ dato che $\sum_{i=1}^{n}(y_i)$ è n volte la media, quindi uguale a $\sum_{i=1}^{n}(\bar{y})$, da questo segue che:

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + \bar{x}\sum_{i=1}^{n}(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}). \tag{1}$$

Analogamente:

$$\sum_{i=1}^{n} x_i(x_i - \bar{x}) = \sum_{i=1}^{n} x_i^2 - \bar{x}x_i = \sum_{i=1}^{n} x_i^2 - 2\bar{x}x_i + \bar{x}^2 + \bar{x}x_i - \bar{x}^2$$

$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 + \bar{x}\sum_{i=1}^{n}(x_i - \bar{x}) \tag{2}$$

dove $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ per lo stesso motivo di sopra, quindi:

$$\sum_{i=1}^{n} x_i(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Sostituendo nell'Equazione 1:

$$\begin{aligned}\sum_{i=1}^{n}(y_i - (ax_i + b)) &= \sum_{i=1}^{n} x_i(y_i - \bar{y}) - a\sum_{i=1}^{n} x_i(x_i - \bar{x}) \text{ per 1) e 2)} \\ &= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) - a\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0.\end{aligned}$$

Isolando a:

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

7. Soluzione finale

I coefficienti ottimali sono:

$$\boxed{a^* = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad b^* = \bar{y} - a^*\bar{x}}.$$

8. Conclusione

Si nota che $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ è la covarianza tra x e y (cov_{xy}) e che $\sum_{i=1}^{n}(x_i - \bar{x})^2$ è la varianza di x (var_x), quindi possiamo riscrivere $a*$ come:

$$a^* = \frac{cov_{xy}}{var_x}.$$

Nota: La formula per a^* è il *coefficiente angolare* della retta di regressione, mentre b^* è l'*intercetta* che assicura che la retta passi per il punto $(\bar{x}, \bar{y})$.