Regressione lineare di un insieme di dati

È dato un insieme $A = \{(x_i;\, y_i),\ i = 1, \ldots, n\}$ di dati sperimentali. Vogliamo determinare la retta di equazione $y = mx + q$ che meglio approssima l'insieme A con il criterio dei minimi quadrati. Ciò significa che vogliamo rendere minima la somma degli scarti quadrati S, con:

$$S = \sum_{i=1}^{n} [y_i - (mx_i + q)]^2.$$

A questo scopo, occorre risolvere il sistema lineare che si ottiene imponendo le condizioni

$$\begin{cases} \dfrac{\partial S}{\partial m} = 0 \\ \dfrac{\partial S}{\partial q} = 0 \end{cases} \tag{1}$$

Calcoliamo quindi

$$\begin{aligned} 0 = \frac{\partial S}{\partial m} &= \sum_{i=1}^{n} \{-2[y_i - (mx_i + q)]x_i\} \\ &= -2 \sum_{i=1}^{n} (y_i x_i - m x_i^2 - q x_i), \end{aligned}$$

da cui otteniamo la condizione:

$$m \sum_{i=1}^{n} x_i^2 + q \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i.$$

Passiamo ora alla seconda equazione

$$\begin{aligned} 0 = \frac{\partial S}{\partial q} &= \sum_{i=1}^{n} \{-2[y_i - (mx_i + q)]\} \\ &= -2 \sum_{i=1}^{n} (y_i - m x_i - q), \end{aligned}$$

fornisce la condizione

$$m \sum_{i=1}^{n} x_i + nq = \sum_{i=1}^{n} y_i.$$

Introducendo le notazioni

$$\Sigma_1 \equiv \sum_{i=1}^{n} x_i, \quad \Sigma_2 \equiv \sum_{i=1}^{n} x_i^2, \quad \Xi_0 \equiv \sum_{i=1}^{n} y_i, \quad \Xi_1 \equiv \sum_{i=1}^{n} x_i y_i,$$

il sistema (1) può essere scritto come

$$\begin{cases} m\Sigma_2 + q\Sigma_1 = \Xi_1 \\ m\Sigma_1 + qn = \Xi_0 \end{cases} \tag{2}$$

Il sistema lineare (2) può essere risolto con metodi standard. La soluzione è:

$$\begin{cases} m = \dfrac{\Xi_0 \Sigma_1 - n\Xi_1}{\Sigma_1^2 - n\Sigma_2} \\ q = \dfrac{\Xi_1 \Sigma_1 - \Xi_0 \Sigma_2}{\Sigma_1^2 - n\Sigma_2} \end{cases} \tag{3}$$