Regressione quadratica di un insieme di dati

È dato un insieme $A = \{(x_i;\, y_i),\ i = 1, \ldots, n\}$ di dati sperimentali. Vogliamo determinare la parabola di equazione $y = ax^2 + bx + c$ 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:

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

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

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \\ \dfrac{\partial S}{\partial c} = 0 \end{cases} \tag{1}$$

Calcoliamo quindi

$$\begin{aligned} 0 = \frac{\partial S}{\partial a} &= \sum_{i=1}^{n} \left\{ -2[y_i - (ax_i^2 + bx_i + c)]x_i^2 \right\} \\ &= -2 \sum_{i=1}^{n} (y_i x_i^2 - a x_i^4 - b x_i^3 - c x_i^2), \end{aligned}$$

da cui otteniamo la prima equazione del sistema:

$$a \sum_{i=1}^{n} x_i^4 + b \sum_{i=1}^{n} x_i^3 + c \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 y_i.$$

Ora passiamo alla seconda equazione:

$$\begin{aligned} 0 = \frac{\partial S}{\partial b} &= \sum_{i=1}^{n} \left\{ -2[y_i - (ax_i^2 + bx_i + c)]x_i \right\} \\ &= -2 \sum_{i=1}^{n} (y_i x_i - a x_i^3 - b x_i^2 - c x_i), \end{aligned}$$

da cui otteniamo la condizione:

$$a \sum_{i=1}^{n} x_i^3 + b \sum_{i=1}^{n} x_i^2 + c \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i.$$

Infine, la terza equazione

$$\begin{aligned} 0 = \frac{\partial S}{\partial c} &= \sum_{i=1}^{n} \left\{ -2[y_i - (ax_i^2 + bx_i + c)] \right\} \\ &= -2 \sum_{i=1}^{n} (y_i - a x_i^2 - b x_i - c), \end{aligned}$$

fornisce la condizione

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + nc = \sum_{i=1}^{n} y_i.$$

Introducendo le notazioni

$$\Sigma_1 \equiv \sum_{i=1}^{n} x_i, \quad \Sigma_2 \equiv \sum_{i=1}^{n} x_i^2, \quad \Sigma_3 \equiv \sum_{i=1}^{n} x_i^3, \quad \Sigma_4 \equiv \sum_{i=1}^{n} x_i^4,$$

$$\Xi_0 \equiv \sum_{i=1}^{n} y_i, \quad \Xi_1 \equiv \sum_{i=1}^{n} x_i y_i, \quad \Xi_2 \equiv \sum_{i=1}^{n} x_i^2 y_i,$$

il sistema (1) può essere scritto come

$$\begin{cases} a\Sigma_4 + b\Sigma_3 + c\Sigma_2 = \Xi_2 \\ a\Sigma_3 + b\Sigma_2 + c\Sigma_1 = \Xi_1 \\ a\Sigma_2 + b\Sigma_1 + cn = \Xi_0 \end{cases} \tag{2}$$

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

$$\begin{cases} a = \dfrac{\Xi_0(\Sigma_2^2 - \Sigma_1\Sigma_3) + \Xi_1(n\Sigma_3 - \Sigma_1\Sigma_2) + \Xi_2(\Sigma_1^2 - n\Sigma_2)}{\Sigma_2^3 + \Sigma_1^2\Sigma_4 + n(\Sigma_3^2 - \Sigma_2\Sigma_4) - 2\Sigma_1\Sigma_2\Sigma_3} \\ b = \dfrac{\Xi_0(\Sigma_1\Sigma_4 - \Sigma_2\Sigma_3) + \Xi_1(\Sigma_2^2 - n\Sigma_4) + \Xi_2(n\Sigma_3 - \Sigma_1\Sigma_2)}{\Sigma_2^3 + \Sigma_1^2\Sigma_4 + n(\Sigma_3^2 - \Sigma_2\Sigma_4) - 2\Sigma_1\Sigma_2\Sigma_3} \\ c = \dfrac{\Xi_0(\Sigma_3^2 - \Sigma_2\Sigma_4) + \Xi_1(\Sigma_1\Sigma_4 - \Sigma_2\Sigma_3) + \Xi_2(\Sigma_2^2 - \Sigma_1\Sigma_3)}{\Sigma_2^3 + \Sigma_1^2\Sigma_4 + n(\Sigma_3^2 - \Sigma_2\Sigma_4) - 2\Sigma_1\Sigma_2\Sigma_3} \end{cases} \tag{3}$$