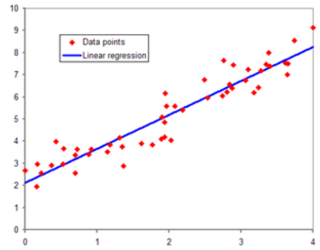
METODO DEI MINIMI QUADRATI

Il *metodo dei minimi quadrati* o *metodo della regressione* è una tecnica di ottimizzazione che permette di trovare una funzione che si avvicini il più possibile ad un'interpolazione di un insieme di dati.



Regressione semplice: una sola variabile in entrata
Regressione multipla: più di una variabile in entrata

REGRESSIONE LINEARE SEMPLICE

$$Y = \beta X + \alpha + \varepsilon$$

Ipotesi:

1. X deterministica, cioè l'errore è così piccolo rispetto ad ε che può essere trascurato
2. Y_k sono variabili casuali gaussiane:
$$Y_k \sim N(\beta X_k + \alpha, \sigma^2)$$

Il nostro obiettivo è cercare i valori degli stimatori B e A (risp. di β e α) che minimizzano la Square Sum:

$$SS = \sum_{k=1}^{n} \big(Y_k - (BX_k + A)\big)^2$$

$$\begin{cases} \dfrac{\partial SS}{\partial A} = 0 \\ \dfrac{\partial SS}{\partial B} = 0 \end{cases} \rightarrow \begin{cases} -2\displaystyle\sum_{k=1}^{n}(Y_k - BX_k - A) = 0 \\ -2\displaystyle\sum_{k=1}^{n} X_k(Y_k - BX_k - A) = 0 \end{cases} \rightarrow \begin{cases} n\bar{Y} - Bn\bar{X} - nA = 0 \\ \displaystyle\sum_{k=1}^{n}(X_kY_k) - B\sum_{k=1}^{n}\left({X_k}^2\right) - An\bar{X} = 0 \end{cases}$$

$$\rightarrow \begin{cases} A = \bar{Y} - B\bar{X} \\ \displaystyle\sum_{k=1}^{n}(X_kY_k) - B\sum_{k=1}^{n}\left({X_k}^2\right) - n\bar{X}\bar{Y} + nB\bar{X}^2 = 0 \end{cases} \rightarrow \begin{cases} \\ B\left(\displaystyle\sum_{k=1}^{n}\left({X_k}^2\right) - n\bar{X}^2\right) = \sum_{k=1}^{n}(X_kY_k) - n\bar{X}\bar{Y} \end{cases}$$

$$\rightarrow \begin{cases} \\ B = \dfrac{\sum_{k=1}^{n} Y_k(X_k - \bar{X})}{\sum_{k=1}^{n}\left({X_k}^2\right) - n\bar{X}^2} \end{cases} \rightarrow \begin{cases} A = \bar{Y} - B\bar{X} \\ B = \dfrac{\sum_{k=1}^{n} Y_k(X_k - \bar{X})}{\sum_{k=1}^{n}\left({X_k}^2\right) - n\bar{X}^2} \end{cases}$$

$$E[B] = E\left[\frac{\sum_{k=1}^{n} Y_k(X_k - \bar{X})}{\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2}\right] = \frac{\sum_{k=1}^{n}(X_k - \bar{X})}{\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2} E[Y_k] = \frac{\sum_{k=1}^{n}(X_k - \bar{X})}{\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2}(\beta X_k + \alpha)$$
$$= \beta\frac{\sum_{k=1}^{n}\left({X_k}^2 - X_k\bar{X}\right)}{\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2} + \alpha\frac{\sum_{k=1}^{n}(X_k - \bar{X})}{\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2} = \beta\frac{\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2}{\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2} + \alpha\frac{n\bar{X} - n\bar{X}}{\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2} = \beta$$

$$E[A] = E[\bar{Y} - B\bar{X}] = E[\bar{Y}] - \bar{X}E[B] = E\left[\sum_{k=1}^{n}\frac{Y_k}{n}\right] - \beta\bar{X} = \frac{1}{n}\sum_{k=1}^{n}E[Y_k] - \beta\bar{X} = \frac{1}{n}\sum_{k=1}^{n}(\beta X_k + \alpha) - \beta\bar{X} = 0 + \frac{n\alpha}{n} = \alpha$$

$$Var(B) = Var\left(\frac{\sum_{k=1}^{n}(X_k - \bar{X})}{\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2} Y_k\right) = Var\left(\sum_{k=1}^{n}\gamma_k Y_k\right) = \sum_{k=1}^{n}{\gamma_k}^2 Var(Y_k)$$
$$= \sigma^2\sum_{k=1}^{n}{\gamma_k}^2 = \sigma^2\frac{\sum_{k=1}^{n}\left({X_k}^2 - \bar{X}^2 - 2X_k\bar{X}\right)}{\left(\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2\right)^2} = \sigma^2\frac{\sum_{k=1}^{n}\left({X_k}^2 - 2X_k\bar{X}\right) + n\bar{X}^2}{\left(\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2\right)^2}$$
$$= \sigma^2\frac{\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2}{\left(\sum_{k=1}^{n} {X_k}^2 - n\bar{X}^2\right)^2} = \frac{\sigma^2}{\sum_{k=1}^{n}(X_k - \bar{X})^2}$$

$$Var(A) = \frac{\sigma^2\sum_{k=1}^{n} {X_k}^2}{n\sum_{k=1}^{n}(X_k - \bar{X})^2}$$