Appunti su “Regressione lineare”

Luigi Glielmo

Corso di Sistemi Dinamici

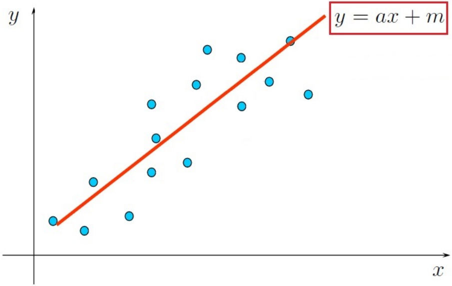
a.a. 2014/2015

Regressione Lineare

- Si supponga di sapere che una variabile dipendente y ed una variabile indipendente x sono legate da una relazione lineare:

$$y = ax + m$$

- Le costanti a e m non sono note, ma si dispone di un numero N di dati sperimentali: (x_i, y_{m_i}) con $i = 1, \ldots, N$.



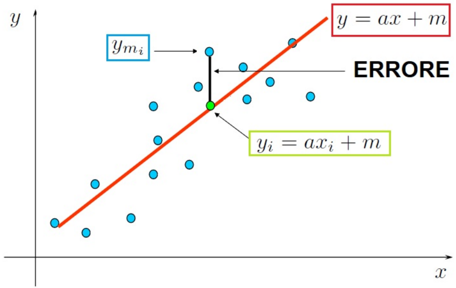
Il problema consiste nel calcolare i parametri a ed m che meglio rappresentano l'insieme delle coppie misurate (x_i, y_{m_i}).

Stima dei parametri a e m

Il metodo generalmente utilizzato per risolvere tale problema è quello della stima ai **minimi quadrati**:

- Secondo tale metodo si cercano i coefficienti a e m che rendono minime le differenze tra i valori misurati y_{m_i} e quelli previsti sulla retta di regressione y_i:

$$\min_{a,m} J(a,m) = \sum_{i=1}^{N} (y_{m_i} - y_i)^2 = \sum_{i=1}^{N} (y_{m_i} - (ax_i + m))^2$$



Stima dei parametri a e m

Si cerca il punto di minimo di $J(a, m)$ ponendo le derivate parziali rispetto ad a e m pari a zero (questa è una condizione necessaria, non sufficiente; una condizione sufficiente richiede anche il calcolo delle derivate seconde, ma la si omette per semplicità). Dunque

$$\frac{\partial J}{\partial a} = 0 \qquad \frac{\partial J}{\partial m} = 0 \qquad \text{(due equazioni in due incognite!)}$$

Derivando $J(a, m)$ rispetto ad a e m si ottiene:

$$\frac{\partial J}{\partial a} = \frac{\partial}{\partial a} \sum_{i=1}^{N} (y_{m_i} - a x_i - m)^2 = \sum_{i=1}^{N} \frac{\partial}{\partial a} (y_{m_i} - a x_i - m)^2 =$$

$$= \sum_{i=1}^{N} 2(y_{m_i} - a x_i - m)(-x_i) = \sum_{i=1}^{N} 2(y_{m_i} x_i - a x_i^2 - m x_i)$$

$$\frac{\partial J}{\partial m} = \frac{\partial}{\partial m} \sum_{i=1}^{N} (y_{m_i} - a x_i - m)^2 = \sum_{i=1}^{N} \frac{\partial}{\partial m} (y_{m_i} - a x_i - m)^2 =$$

$$= \sum^{N} 2(y_{m_i} - a x_i - m)(-1) = -\sum^{N} 2(y_{m_i} - a x_i - m)$$

Stima dei parametri a e m

Le soluzioni volute per i parametri a e m si ottengono risolvendo il sistema di equazioni (lineari!):

$$\begin{cases} \frac{\partial J}{\partial a} = 0 \quad \rightarrow \quad a\sum_{i=1}^{N} x_i^2 + m\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_{m_i} x_i \\ \frac{\partial J}{\partial m} = 0 \quad \rightarrow \quad a\sum_{i=1}^{N} x_i + mN = \sum_{i=1}^{N} y_{m_i} \end{cases}$$

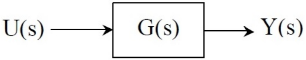
Si ha:

$$a = \frac{N\sum_{i=1}^{N} x_i y_{m_i} - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_{m_i}}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}$$

$$m = \frac{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_{m_i} - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_{m_i}}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}$$

Un'applicazione

Stima del guadagno statico di un sistema



- Il valore dell'uscita a regime di un sistema asintoticamente stabile in risposta ad un ingresso costante è dato dalla relazione:

$$\bar{y} = \mu \bar{u}$$

  dove μ rappresenta il guadagno statico del sistema.

- è possibile calcolare μ a partire da un insieme di misure dell'uscita *a regime* rispetto ad un ingresso costante $(\bar{u}_i, \bar{y}_{m_i})$; applicando il metodo dei minimi quadrati sopra visto (ma nel caso particolare di intercetta $m = 0$, attenzione: abbiamo una sola incognita, non due), si ha

$$\mu = \frac{\sum_{i=1}^{N} (\bar{y}_{m_i} \bar{u}_i)}{\sum_{i=1}^{N} \bar{u}_i^2}$$