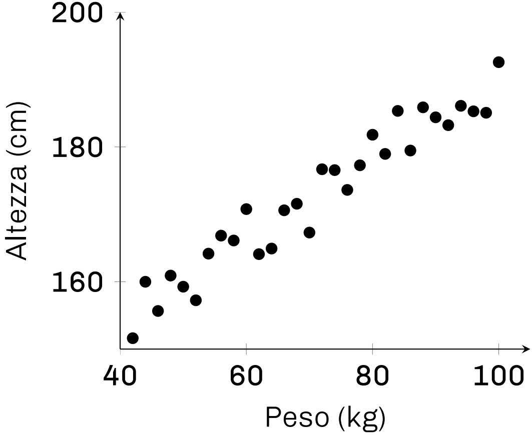
Algoritmi Numerici (Parte IV)

Lezione 4: regressione lineare

Alessandro Antonucci
alessandro.antonucci@supsi.ch

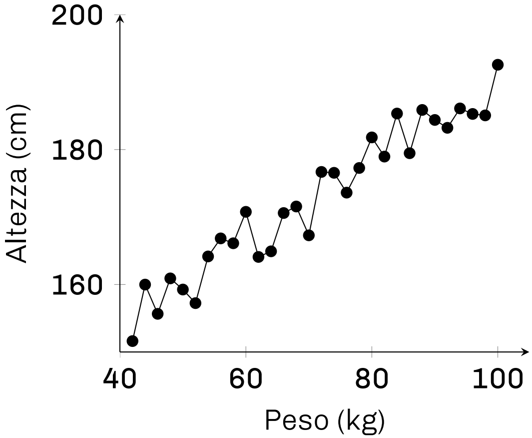
Motivazioni

- Dati su peso e altezza di 30 persone
- Interpolazione poco significativa
- Regressione (lineare) per estrapolare conoscenza dai dati



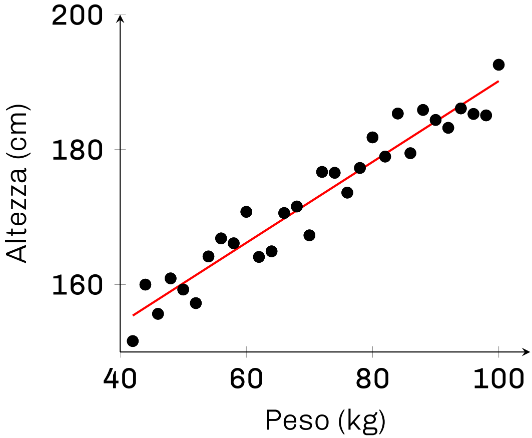
Motivazioni

- Dati su peso e altezza di 30 persone
- Interpolazione poco significativa
- Regressione (lineare) per estrapolare conoscenza dai dati



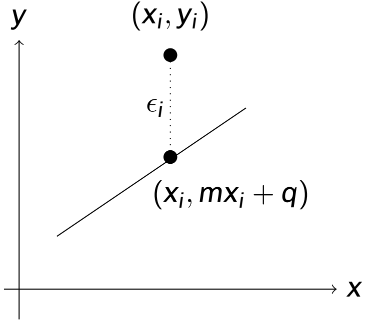
Motivazioni

- Dati su peso e altezza di 30 persone
- Interpolazione poco significativa
- Regressione (lineare) per estrapolare conoscenza dai dati



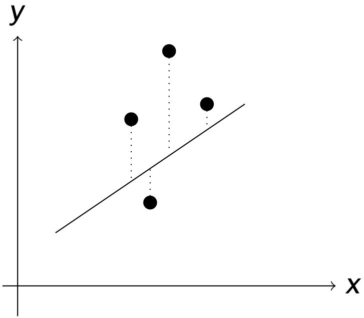
Regressione lineare

- Dati n punti $\{(x_i, y_i)\}_{i=1}^n$, cerco una funzione che descriva la relazione fra x e y
- Relazione lineare: $y = mx + q$, ma come trovare m e q?
- Distanza sulla verticale fra punto e retta $\epsilon_i := y_i - mx_i - q$



Regressione lineare

- Minimizzare $\sum_i \epsilon_i$?
- No, $\epsilon_i < 0$ se punto sotto retta
- Meglio minimizzare $\sum_i \epsilon_i^2$
- Ottimizzazione funzione di due variabili (m e q)



$$F(m, q) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - mx_i - q)^2$$

Ottimizzazione

- Minimizzare/massimizzare $f(x)$? Pongo $\frac{\mathrm{d}f}{\mathrm{d}x} = 0$
- Minimizzare/massimizzare $F(m, q)$? $\frac{\mathrm{d}F}{\mathrm{d}m} = \frac{\mathrm{d}F}{\mathrm{d}q} = 0$

$$\begin{cases} \frac{\mathrm{d}F}{\mathrm{d}q} &= -2\sum_i (y_i - mx_i - q) &= 0 \\ \frac{\mathrm{d}F}{\mathrm{d}m} &= -2\sum_i x_i(y_i - mx_i - q) &= 0 \end{cases}$$

$$\begin{cases} n\overline{y} - nm\overline{x} - nq = 0 \\ \sum_i x_i y_i - m\sum_i x_i^2 - q\sum_i x_i = 0 \end{cases}$$

$$\begin{cases} q = \overline{y} - m\overline{x} \\ \sum_i x_i y_i - m\sum_i x_i^2 = n\overline{x}(\overline{y} - m\overline{x}) \end{cases}$$

$$\begin{cases} q = \overline{y} - m\overline{x} \\ \sum_i x_i y_i - n\overline{x}\cdot\overline{y} = m(\sum_i x_i^2 - n\overline{x}^2) \end{cases}$$

con $\overline{x} = \frac{\sum_i x_i}{n}$ e $\overline{y} = \frac{\sum_i y_i}{n}$

Regressione Lineare (formule)

La retta è $y = mx + q$

con

$$\begin{cases} q = \overline{y} - m\overline{x} \\ m = \frac{\sum_i x_i y_i - n\overline{x} \cdot \overline{y}}{\sum_i x_i^2 - n\overline{x}^2} \end{cases}$$

e

$\overline{x} = \frac{\sum_i x_i}{n}$ e $\overline{y} = \frac{\sum_i y_i}{n}$