



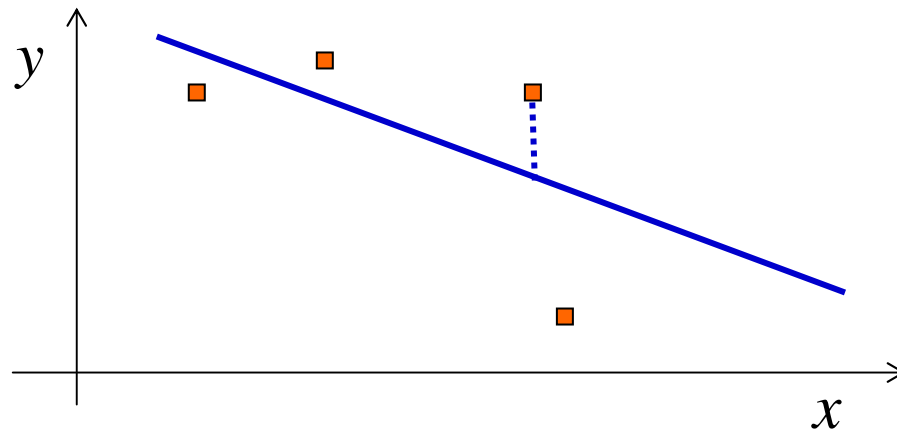
Claudio Arbib
Università di L'Aquila
Ricerca Operativa

Esercitazione IV

Acquistare l'auto ideale
– la soluzione –

Il problema

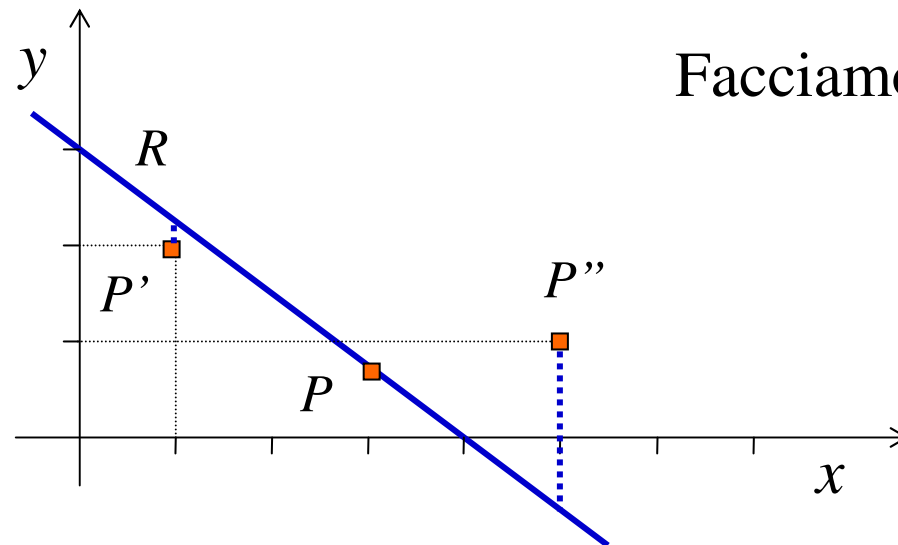
- Come **distanza di una retta da un insieme di punti** abbiamo assunto la **somma delle distanze** di ciascun punto dalla retta
- Abbiamo inoltre scelto di definire la **distanza di un punto** (x_i, y_i) da una retta come la differenza tra l'ordinata y_i del punto e quella della retta nel punto x_i .



Problema:

Qual è la retta di regressione associata alla nuvola di punti?

- Una retta nel piano può essere descritta da un'equazione del tipo $y = a_1x + a_0$.
- Ad esempio la retta R disegnata qui sotto, che passa per i punti $(0, 3)$ e $(4, 0)$, ha equazione $y = -\frac{3}{4}x + 3$.



Facciamo comparire dei
punti nel piano

- Il punto $P = (3, \frac{3}{4})$ ha distanza da R pari a 0
- Il punto $P' = (1, 2)$ ha distanza da R pari a $(-\frac{3}{4} \cdot 1 + 3) - 2 = \frac{1}{4}$
- Il punto $P'' = (5, 1)$ ha distanza da R pari a $1 - (-\frac{3}{4} \cdot 5 + 3) = 1 + \frac{3}{4} = \frac{7}{4}$

In generale, il punto (x_i, y_i) avrà distanza $d_i = |y_i - (a_1x_i + a_0)|$

Come calcolare la retta di minima distanza

- In sostanza, la distanza d_i del punto (x_i, y_i) da R dipende da a_0 e a_1 : infatti $d_i = |y_i - (a_1 x_i + a_0)|$
- Da questi due parametri dipende anche la scelta della retta: infatti noti a_0 e a_1 la retta è completamente definita
- Perciò la retta di minima distanza dalla nuvola di punti (x_1, y_1) , (x_2, y_2) , ..., (x_n, y_n) sarà quella retta definita da due parametri a_0 e a_1 tali da minimizzare la somma $d_1 + d_2 + \dots + d_n$ delle distanze dei punti.

Come calcolare la retta di minima distanza

Il problema da risolvere è dunque

$$\begin{aligned}\min \quad & d_1 + d_2 + \dots + d_n = \\ & = |y_1 - (a_1x_1 + a_0)| + |y_2 - (a_1x_2 + a_0)| + \dots + |y_n - (a_1x_n + a_0)|\end{aligned}$$

Questo è un problema di programmazione **non lineare**

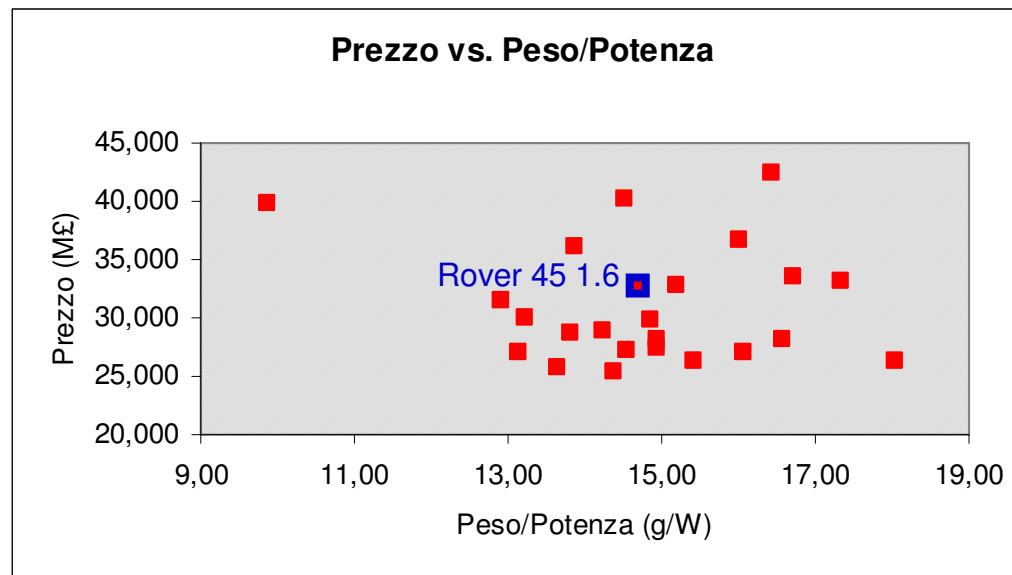
Ma si può rimediare: il problema che risolveremo sarà

\min	$d_1 + d_2 + \dots + d_n$	
con	$d_1 \geq y_1 - (a_1x_1 + a_0)$	$d_1 \geq (a_1x_1 + a_0) - y_1$
	$d_2 \geq y_2 - (a_1x_2 + a_0)$	$d_2 \geq (a_1x_2 + a_0) - y_2$
	\dots	\dots
	$d_n \geq y_n - (a_1x_n + a_0)$	$d_n \geq (a_1x_n + a_0) - y_n$
	$d_1, d_2, \dots, d_n \geq 0$	

in rosso sono indicate le **variabili**

Tornando a noi ...

- Impostiamo il problema in modo da calcolare la retta di regressione relativa al grafico **prezzo** vs. **peso/potenza**, e capire così come si posiziona la *Rover 45*



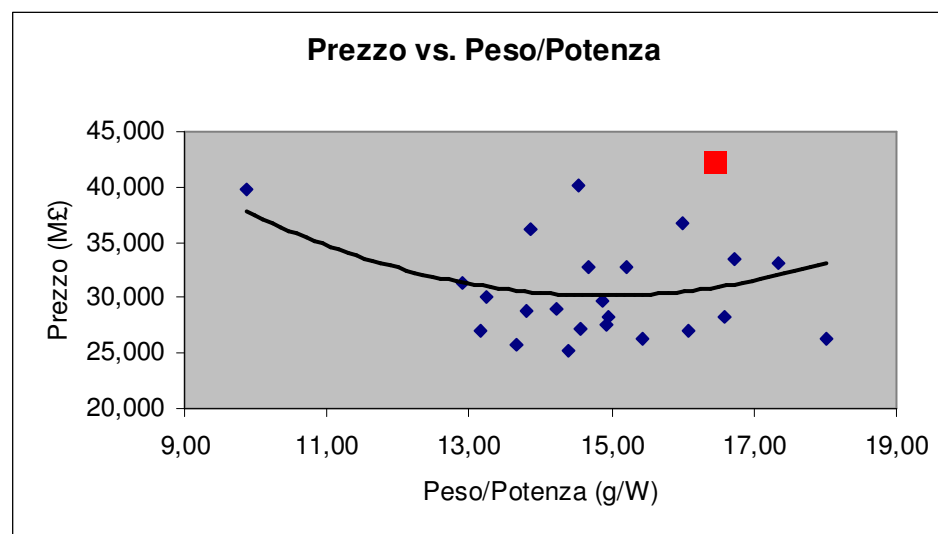
- Possiamo farlo a partire dai dati del foglio Excel [Auto.xls](#)

Altre curve di tendenza

- Finora abbiamo adottato come distanza di un punto $P = (x_1, y_1)$ da una retta R la differenza tra l'ordinata y_1 di P e l'ordinata del punto di R di ascissa x_1 , in valore assoluto (questo corrisponde a immaginare la x_1 non affetta da errore). Altre distanze, come quella euclidea, danno luogo a rette diverse.
- In generale una curva di tendenza può avere **forme diverse** da una retta. Excel offre la possibilità di calcolare curve di tendenza lineari, polinomiali, logaritmiche, etc.
- Per far ciò, dopo aver prodotto il grafico a partire da un insieme di dati rappresentanti coppie di punti, occorre:
 - a) selezionare il grafico usando il tasto destro
 - b) applicare dal menu a tendina il comando “Aggiungi linea di tendenza”
 - c) scegliere il tipo di curva di tendenza desiderato

Altre curve di tendenza

- Qui sotto è evidenziata una linea di tendenza polinomiale del secondo ordine



Quando le qualità sono molteplici

- Se vogliamo riferire il prezzo **a più d'una qualità** (ad es., rapporto peso/potenza e consumi) non possiamo più ricorrere ad Excel: si può infatti **rappresentare** la nuvola di punti per 2 qualità, ma non per 3 o più.
- Inoltre non potremo più parlare di **curva di regressione**. Ad es., per 3 parametri (2 qualità + il prezzo) avremo a che fare con una **superficie di regressione**.
- Se questa superficie esprime una dipendenza lineare, sarà un **piano**. In geometria analitica, un piano è rappresentato da una **equazione** della forma $z = a_0 + a_1x + a_2y$.
- La distanza del punto (x_i, y_i, z_i) dal piano sarà ora pari a
$$d_i = |z_i - (a_0 + a_1x_i + a_2y_i)|$$
e il modello di PL descritto può facilmente essere esteso.