Regresión lineal

Recta que minimiza la distancia entre los puntos observados y los puntos sobre la recta

Residuo → distancia entre un punto observado y el correspondiente punto sobre la recta

$r_i = y_i - \hat{y}_i$ donde $\hat{y}_i$ es el punto iesimo sobre la recta

$$\hat{y}_i = c_1 x_i + c_o$$

$$r_i = y_i - (c_1 x_i + c_o)$$

$$SSR = \sum_{i=1}^{N} r_i^2 = \sum_{i=1}^{N} [y_i - (c_1 x_i + c_o)]^2$$

$$\frac{d\,SSR}{d\,c_o} = \sum_{i=1}^{N} (-2)[y_i - (c_1 x_i + c_o)]$$

$$\frac{d\,SSR}{d\,c_1} = \sum_{i=1}^{N} (-2\,x_i)[y_i - (c_1 x_i + c_o)]$$

$$\sum_{i=1}^{N}(-2)[y_i-(c_1 x_i+c_o)]=0$$

$$\sum_{i=1}^{N}(-2x_i)[y_i-(c_1 x_i+c_o)]=0$$

$$\sum_{i=1}^{N} c_o+\sum_{i=1}^{N} c_1 x_i=\sum_{i=1}^{N} y_i \quad \text{(I)}$$

$$\sum_{i=1}^{N} c_o x_i+\sum_{i=1}^{N} c_1 x_i^2=\sum_{i=1}^{N} x_i y_i \quad \text{(II)}$$

$$c_o N+c_1\sum_{i=1}^{N} x_i=\sum_{i=1}^{N} y_i \quad \text{(I)}$$

$$c_o\sum_{i=1}^{N} x_i+c_1\sum_{i=1}^{N} x_i^2=\sum_{i=1}^{N} x_i y_i \quad \text{(II)}$$

$$c_o+c_1\left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i=\left(\frac{1}{N}\right)\sum_{i=1}^{N} y_i \quad \text{(I)}$$

$$c_o\left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i+c_1\left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i^2=\left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i y_i \quad \text{(II)}$$

Si $\bar{x}=\left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i$ y $\bar{y}=\left(\frac{1}{N}\right)\sum_{i=1}^{N} y_i$

$$c_o+c_1\bar{x}=\bar{y} \quad \text{(I)}$$

$$c_o\left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i+c_1\left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i^2=\left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i y_i \quad \text{(II)}$$

$$\begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i^2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \left(\frac{1}{N}\right)\sum_{i=1}^{N} x_i y_i \end{bmatrix}$$

A x = b

Coeficiente de determinación

$$R^2=1-\frac{\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)}{\sum_{i=1}^{N}\left(y_i-\bar{y}\right)}$$

Código clase

```
import matplotlib.pyplot as plt
import numpy as np

N=100
x=np.sort(np.random.rand(N) * 100)
y=np.sort(np.random.rand(N) * 100)

A=[[1,(1/N)*np.sum(x)],[np.average(x),np.average(x**2)]]
b=[[np.average(y)],[np.average(x*y)]]
c=np.linalg.solve(A,b)
print(c)

p=np.polyfit(x,y,1)
xr=x #np.arange(x[0], x[-1], 1)
yr=np.polyval(p,xr)
print(p)

plt.figure()
plt.plot(x,y, ".r")
plt.plot(xr,yr,"b")
plt.plot(xr,yr,".g")
```