9 Least Square

Friday, March 02, 2012
11:33 AM

Earlier in our class

Basic statistics ...

Suppose we have n observations : $x_1, x_2, x_3, x_4, \ldots, x_n$

Find "b" to represent this data set.

To determine how good "b" is,
we will calculate

$$\sum_{i=1}^{n} (x_i - b)^2$$

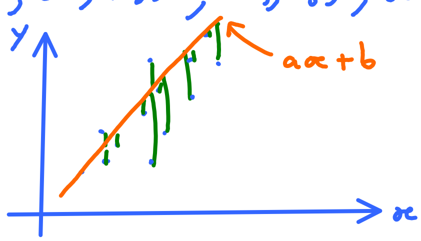
and minimize it.

$$\xrightarrow[\text{calculus}]{} \text{ the best "b"} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\frac{d}{db} = 0$$

Suppose that we have n pairs of observations

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$$



Find the "best" straight line that "fits" the observations.

Again, want to minimize square error

$$\sum_{i=1}^{n} \left(y_i - (a x_i + b) \right)^2$$

EX $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$\longrightarrow$ $\overline{xy} - \bar{x}\,\bar{y}$

calculus

$\frac{\partial}{\partial a} = 0$

$\frac{\partial}{\partial b} = 0$

$$a = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - (\bar{x})^2} \qquad \mathbb{E}Y \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$b = \bar{y} - a\bar{x} \qquad \mathbb{E}[X^2] \quad \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$\underbrace{\mathbb{E}[XY]}_{\text{correlation}} \quad \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$

$$a = \frac{\mathbb{E}[XY] - \mathbb{E}X\,\mathbb{E}Y}{\mathbb{E}[X^2] - (\mathbb{E}X)^2} = \frac{\mathrm{cov}(X,Y)}{\sigma_X^2} = \rho\frac{\sigma_Y}{\sigma_X}$$

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y} \Rightarrow \mathrm{cov}(X,Y) = \rho\,\sigma_X\sigma_Y$$

$$y \approx ax + b$$

$$Y \approx aX + b = aX + (\mathbb{E}Y - a\,\mathbb{E}X)$$

$$= \mathbb{E}Y + a(X - \mathbb{E}X)$$

$$= \mathbb{E}Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mathbb{E}X)$$

Here is how things get crazy ... in statistics ...

$$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \ldots, (X_n, Y_n)$$