Thursday, November 30, 2017 7:46 AM

Lab 13

In this Lab:

- simple linear regression
- multiple linear regression

Simple Linear Regression

Problem: a pair of correlated rvs (X, Y)

we want to estimate Y based on observation X.

Solution:

estimator: $\mathbb{E}[Y|X] = g(X) \to$ a function of X

where

$$g(u) = \mathbb{E}[Y|X=u] = \int v \, f_{Y|X}(v|u) \, dv$$

Thursday, November 30, 2017 8:04 AM

linear estimator:

$$\hat{\mathbb{E}}[Y|X] = L^*(X)$$

where

$$L^*(u) = \mu_Y + \frac{\text{cov}(Y,X)}{\text{Var}(X)}(u - \mu_X)$$

Calculating μ_X, μ_Y, $\text{var}(X)$ and $\text{cov}(Y,X)$ requires knowing the distributions of X, Y and their joint distribution, which often is not available to us.

Often we have some samples (x_1, y_1), (x_2, y_2), $\vdots$, (x_n, y_n)

$$\mu_X \approx \frac{x_1 + \cdots + x_n}{n}$$

$$\mu_Y \approx \frac{y_1 + \cdots + y_n}{n}$$

$$\text{Var}(X) = \mathbb{E}[(X-\mu_X)^2] \approx \frac{(x_1-\mu_X)^2 + \cdots + (x_n - \mu_X)^2}{n}$$

$$\text{cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \approx \frac{(x_1-\mu_X)(y_1-\mu_Y) + \cdots + (x_n-\mu_X)(y_n-\mu_Y)}{n}$$

Another name: linear regression
(of Y values based on X values)

Example 1

X_1 = mid-term1 score Y = Final exam score

Example 2

X_2 = mid-term 2 score Y = Final exam score

Example 3

X_3 = quiz score Y = Final exam score

Think:
We can estimate Y based on X_1, or X_2, or X_3.
But would we be better off by estimating based on all X_1, X_2 and X_3?

Multiple Linear Regression

Problem: same as simple linear regression, but X is a vector instead of a scalar.
↓ dim - n

Before:

$$\hat{\mathbb{E}}[Y|X] = \mathbb{E}[Y] + \frac{\mathrm{Cov}(Y,X)}{\mathrm{Var}(X)}(X - \mathbb{E}[X])$$

$$= \mathbb{E}[Y] + \mathrm{cov}(Y,X)\,\mathrm{cov}(X,X)^{-1}(X - \mathbb{E}[X])$$

Now

$$\hat{\mathbb{E}}[Y/X] = \underset{1\times 1}{\mathbb{E}[Y]} + \underset{1\times n}{\mathrm{cov}(Y,X)}\,\underset{n\times n}{\mathrm{cov}(X,X)^{-1}}\,(\underset{n\times 1}{X} - \underset{n\times 1}{\mathbb{E}[X]})$$

$$\mathrm{cov}(Y,X) = [\mathrm{cov}(Y,X_1), \cdots, \mathrm{cov}(Y,X_n)]$$

$$\mathrm{cov}(X,X) = \begin{bmatrix} \mathrm{cov}(X_1,X_1) & \cdots & \mathrm{cov}(X_1,X_n) \\ \mathrm{cov}(X_2,X_1) & \cdots & \mathrm{cov}(X_2,X_n) \\ \vdots & & \\ \mathrm{cov}(X_n,X_1) & \cdots & \mathrm{cov}(X_n,X_n) \end{bmatrix}$$

Thursday, November 30, 2017 8:28 AM

$\hat{E}[Y | X_1, X_2, X_3]$ should be in the form:

$\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \beta$