Lecture 2: Linear Regression

Instructor: Prof. Shuai Huang

Industrial and Systems Engineering

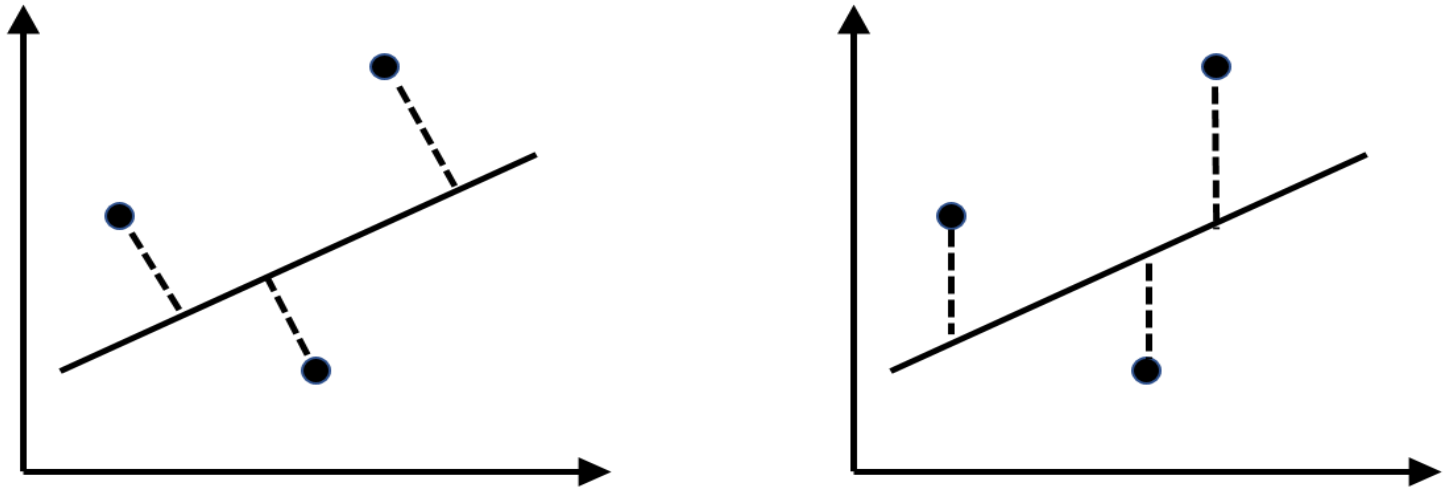
University of Washington

The linear regression model

A simple example: $f(x) = \beta_0 + \beta_1 x$, $\epsilon \sim N(0, \sigma_\varepsilon^2)$

- The linear relationship remains the same for all the values of x (a global pattern).
- The model suggests a fundamental unpredictability of y, if y is generated by a combination of the signal (the $f(x)$) and the noise (ϵ).
- ***R-squared:*** $\frac{\sigma_y^2 - \sigma_\varepsilon^2}{\sigma_y^2}$.
- The noise is usually modeled as gaussian distribution, but this assumption could be relaxed.

Parameter estimation



Two principles to fit a linear regression model: (left) perpendicular offsets; (right) vertical offsets.

In the historic development of linear regression, the paradigm of vertical offsets gained popularity which led to the least-squares estimation

Derivation of the least-squares estimation

- Suppose that we have collected N data points, denoted as, (x_n, y_n) for $n = 1, 2, \ldots, N$.
- The sum of the squared of the vertical derivations of the observed data points from the line is:

$$l(\beta_0, \beta_1) = \sum_{n=1}^{N} [y_n - (\beta_0 + \beta_1 x_n)]^2.$$

- To estimate β_0 and β_1 is to minimize this least-square loss function $l(\beta_0, \beta_1)$.

A simple example

Table 2.2: An exemplary dataset

X	1	3	3	5	5	6	8	9
Y	2	3	5	4	6	5	7	8

The R-code to verify your calculation:

```
## Simple example of regression with one predictor
data = data.frame(rbind(c(1,2),c(3,3),c(3,5),c(5,4),c(
5,6),c(6,5),c(8,7),c(9,8)))
colnames(data) = c("Y","X")
str(data)

lm.YX <- lm(Y ~ X, data = data)
summary(lm.YX)
```

Extension to multiple linear regression

- There are more than one predictor: $y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \varepsilon$
- In matrix form: $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$,

$$\text{where } \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1N} & x_{2N} & \cdots & x_{pN} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix},$$

$$\text{and } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}.$$

Derivation of the least-squares estimation

To estimate $\boldsymbol{\beta}$, we can derive the optimization formulation in matrix form as:

$$\min_{\boldsymbol{\beta}} (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Take the gradient of the objective function and set it to be zero:

$$\frac{\partial (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0.$$

This leads to the least square estimator of $\boldsymbol{\beta}$ as

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{Y}.$$

Notice the resemblance between $\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{Y}$ with $\beta_1 = \frac{cov(x,y)}{var(x)}$!

Hypothesis testing of regression parameters

- It is important to recognize that, since $\boldsymbol{y}$ is a random vector and induce uncertainty, $\widehat{\boldsymbol{\beta}}$ is a random vector as well.
- The mean of $\widehat{\boldsymbol{\beta}}$ is $\boldsymbol{\beta}$, as

$$E(\widehat{\boldsymbol{\beta}}) = E\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{y}\right] = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T E[\boldsymbol{y}] = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}.$$

- The covariance matrix of $\widehat{\boldsymbol{\beta}}$ can be readily derived as

$$cov(\widehat{\boldsymbol{\beta}}) = \sigma_\varepsilon^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}.$$

- Thus, it is readily available to derive the hypothesis testing procedure for any regression parameter.

R lab

- Download the markdown code from course website
- Conduct the experiments
- Interpret the results
- Repeat the analysis on other datasets