MULTIVARIATE STATISTICAL ANALYSIS

Lecture 3
Multivariate Linear Regression

Associate Professor Lý Quốc Ngọc

TRUONG DAI HOC KHOA HOC TU NHIEN TP. HO CHI MINH

KHTN

cdio 4.0

KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

fit@hcmus

Contents

3.1. Purpose of LR

Regression analysis is the statistical methodology for predicting values of one or more response (**dependent**) variables from a collect of predictor (**independent**) variable values..

3.2. Problem statement

Let $z_1, z_2, \ldots, z_r$ be r predictor variables to be related to a response variable Y .
The linear regression model with a single response takes the form:

$$Y = \beta_0 + \beta_1 z_1 + \ldots + \beta_r z_r + \varepsilon$$

3.2. Problem statement

Let $z_{i1}, z_{i2}, \ldots, z_{ir}$ be r predictor variables to be related to a response variable $Y_i, i = 1..n$

The linear regression model with n response takes the form:

$$Y_1 = \beta_0 + \beta_1 z_{11} + \ldots + \beta_r z_{1r} + \varepsilon_1$$

$$Y_2 = \beta_0 + \beta_1 z_{21} + \ldots + \beta_r z_{2r} + \varepsilon_2$$

.

.

$$Y_n = \beta_0 + \beta_1 z_{n1} + \ldots + \beta_r z_{nr} + \varepsilon_n$$

Where ε_i are assumed:

$$1.\, E(\varepsilon_i) = 0;\ 2.\, Var(\varepsilon_i) = \sigma^2;\ 3.\, Cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j$$

fit@hcmus

3.2. Problem statement

Let $z_{i1}, z_{i2}, \ldots, z_{ir}$ be r predictor variables to be related to a response variable $Y_i, i = 1..n$

The linear regression model with n response in the form of the matrix:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{12} \cdots & z_{1r} \\ 1 & z_{21} & z_{22} \cdots & z_{2r} \\ \vdots & \vdots & & \vdots \\ 1 & z_{n1} & z_{n2} \cdots & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$Y = Z.\beta + \varepsilon$$

ε are assumed: $1.\ E(\varepsilon) = 0;\ 2.\ Cov(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I$

3.3. Method

Let $z_1, z_2, \dots, z_r$ be r predictor variables to be related to a response variable Y .
The linear regression model with a single response takes the form:

$$Y = \beta_0 + \beta_1 z_1 + \dots + \beta_r z_r + \varepsilon$$

Estimate $\{\beta_i\}, i = 1..r$ based on the samples having the corresponding between $\{z_{jr}\}, j = 1..n$ and $\{y_j\}, j = 1..n$

Estimate $\{\beta_i\}, i = 1..r$ to minimize sum of squares of the differences between the predicting values and groundtruth values.

3.3. Phương pháp

Sum of squares of the differences between the prediting values and groundtruth values :

$$S(\beta) = \sum_{j=1}^{n} (y_i - \beta_0 - \beta_1 z_{j1} - \ldots - \beta_r z_{jr})^2 = (y - Z\beta)^T.(y - Z\beta)$$

The necessary condition for the above quantity to be minimized is:

$$\frac{\partial S(\beta)}{\partial \beta} = 0$$

$$\Rightarrow -2\frac{\partial(\beta^T Z^T y)}{\partial \beta} + \frac{\partial(\beta^T Z^T Z\beta)}{\partial \beta} = 0$$

$$\Rightarrow -2Z^T y + 2Z^T Z\beta = 0$$

$$\Rightarrow \beta = (Z^T Z)^{-1} Z^T y$$

3.4. Geometrical Explanation

Discuss in class

3.5. Model Checking

Discuss in class

3.6. Case study

Discuss in class