KHOA CÔNG NGHỆ THÔNG TIN-ĐHKHTN

PHÂN TÍCH THỐNG KÊ DỮ LIỆU NHIỀU BIẾN

Giảng viên: PGS.TS. Lý Quốc Ngọc

TPHCM, 8-2020

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

fit@hcmus

PHÂN TÍCH THỐNG KÊ DỮ LIỆU NHIỀU BIẾN

Bài giảng 3:
Mô hình hồi quy tuyến tính nhiều biến

Giảng viên: PGS.TS. Lý Quốc Ngọc

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

fit@hcmus

Nội dung

3. Mô hình hồi quy tuyến tính nhiều biến

3.1. Mục đích

3.2. Phát biểu bài toán

3.3. Phương pháp

3.4. Ý nghĩ hình học

3.5. Kiểm chứng tính đúng đắn của mô hình

3.6. Ví dụ

3.1. Mục đích

Phân tích hồi quy là phương pháp thống kê để tiên đoán của một hay nhiều biến phụ thuộc từ tập biến độc lập.

3.2. Phát biểu bài toán

Giả sử $z_1, z_2, \ldots, z_r$ là r biến độc lập có liên quan đến biến phụ thuộc Y

Mô hình hồi quy tuyến tính với một biến kết quả tiên đoán có dạng:

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

fit@hcmus

3.2. Phát biểu bài toán

Giả sử $z_{i1}, z_{i2}, ..., z_{ir}$ là biến độc lập có liên quan đến biến phụ thuộc $Y_i, i = 1..n$

Mô hình hồi quy tuyến tính với n biến kết quả tiên đoán có dạng:

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

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

$$.$$

$$.$$

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

Giả thiết ε_i thỏa: $1.E(\varepsilon_i) = 0; 2.Var(\varepsilon_i) = \sigma^2; Cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j$

3.2. Phát biểu bài toán

Giả sử $z_{i1}, z_{i2}, \ldots, z_{ir}$ là r biến độc lập có liên quan đến biến phụ thuộc $Y_i, i = 1..n$

Mô hình hồi quy tuyến tính với n biến kết quả tiên đoán ở dạng ma trận:

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

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

Giả thiết ε thỏa: $1.E(\varepsilon) = 0;\ 2.Cov(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I$

3.3. Phương pháp

Giả sử $z_1, z_2, ..., z_r$ là r biến độc lập có liên quan đến biến phụ thuộc Y

Mô hình hồi quy tuyến tính với một biến kết quả tiên đoán có dạng:

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

Cần tìm $\{\beta_i\}, i = 1..r$ với tập giá trị thử nghiệm $\{z_{jr}\}, j = 1..n$

ứng với $\{y_j\}, j = 1..n$

3.3. Phương pháp

Xét tổng bình phương độ lệch giữa giá trị thực và giá trị tiên đoán của mô hình:

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

Điều kiện cần để đại lượng trên đạt cực tiểu là:

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

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

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

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

3.4. Ý nghĩa hình học

Bàn luận trên lớp

3.5. Kiểm chứng tính đúng đắn của mô hình

Bàn luận trên lớp

3.6. Ví dụ

Bàn luận trên lớp