Stat 564
10-14-14

1

The general linear model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

Have n observations

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1$$
$$\vdots$$
$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n$$

2

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}_{n\times(k+1)} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$Y = X \beta + \varepsilon$$

X: design matrix

β: parameter vector

③

Still want to minimize

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

$$Y - \hat{Y} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}$$

④

$$(Y - \hat{Y})'(Y - \hat{Y})$$

$$= \left[(y_1 - \hat{y}_1) \quad (y_2 - \hat{y}_2) \quad \cdots \quad (y_n - \hat{y}_n) \right] \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}$$

$$= SSE$$

Also, $\hat{Y} = X\hat{\beta}$

So $SSE = (Y - X\hat{\beta})'(Y - X\hat{\beta})$

Special case 1: (linear) $\dfrac{\partial(\vec{a}'\vec{x})}{\partial(\vec{x})} = \dfrac{\partial\left(\sum_{i=1}^{p} a_i x_i\right)}{\partial(\vec{x})}$

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \vec{a}$$

assume symmetric

Special case 2: (quadratic) $\dfrac{\partial(\vec{x}'A\vec{x})}{\partial\vec{x}}$

$$= \frac{\partial}{\partial\vec{x}}\left[[x_1, x_2, \ldots, x_p]\begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & & \\ a_{p1} & \cdots & a_{pp} \end{bmatrix}_{p\times p}\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}_{p\times 1}\right]$$

$$= \frac{\partial}{\partial\vec{x}}\left[\underset{1\times p}{[x_1, x_2, \ldots, x_p]}\underset{p\times 1}{\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p \\ \vdots \\ a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pp}x_p \end{bmatrix}}\right]$$

$$= \frac{\partial}{\partial\vec{x}}\left[x_1(a_{11}x_1 + \ldots + a_{1p}x_p) + \ldots + x_p(a_{p1}x_1 + \ldots + a_{pp}x_p)\right]$$

$$= \frac{\partial}{\partial\vec{x}}\left[\sum_{i=1}^{p} a_{ii}x_i^2 + 2\sum\sum_{i<j} a_{ij}x_i x_j\right]$$

$$= \begin{bmatrix} 2a_{11}x_1 + 2a_{12}x_2 + 2a_{13}x_3 + \ldots + 2a_{1p}x_p \\ \vdots \\ 2a_{pp}x_p + 2a_{1p}x_1 + \ldots + 2a_{p(p-1)}x_{p-1} \end{bmatrix}$$

7

$$SSE = (Y' - \hat{\beta}'X')(Y - X\hat{\beta})$$

$$= \underset{1\times n}{Y'}\underset{n\times 1}{Y} - \underset{1\times n}{Y'}\underset{n\times(k+1)}{X}\underset{(k+1)\times 1}{\hat{\beta}} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

$$= Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

Review: If f is a real-valued function,

then $\dfrac{\partial f}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$

8

$$= 2\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ \vdots & & & \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

$$= 2A\vec{x}$$

$$SSE = Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

$$\frac{\partial SSE}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} \overset{\text{set}}{=} 0$$

$$X'X\hat{\beta} = X'Y$$

(9)

$$\hat{\beta} = (X'X)^{-1} X'Y \quad \left[\text{Assuming } (X'X)^{-1} \text{ exists}\right]$$

This is the $\hat{\beta}$ that minimizes SSE.

Now $SSE = Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$

$$= Y'Y - 2Y'X(X'X)^{-1}X'Y + Y'X(X'X)^{-1}\cancel{(X'X)}\cancel{(X'X)^{-1}}X'Y$$

$$= Y'Y - Y'X(X'X)^{-1}X'Y$$

(10)

$$SSE = Y'(I - \underbrace{X(X'X)^{-1}X'}_{H})Y$$

"hat" matrix

$$SSE = Y'(I - H)Y$$

Why is H called the "hat" matrix?

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = HY$$