Least Squares Estimation: Continued I

Using matrix representation, we can express the regression model as

$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times(p+1)}\boldsymbol{\beta}_{(p+1)\times 1} + \mathbf{e}_{n\times 1}.$$

The least squares method estimates $\boldsymbol{\beta}$ by minimizing

$$\begin{aligned} \mathrm{RSS}(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\Big(y_i - \beta_0 - x_{i1}\beta_1 - \cdots - x_{ip}\beta_p\Big)^2 \\ &= \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2. \end{aligned}$$

Least Squares Estimation: Continued II

Differentiating RSS($\boldsymbol{\beta}$) with respect to $\boldsymbol{\beta}$ and setting to zero, we have

$$\frac{\partial \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2}{\partial \boldsymbol{\beta}} = \mathbf{0}_{(p+1)\times 1}$$

Least Squares Estimation: Continued II

Differentiating RSS($\boldsymbol{\beta}$) with respect to $\boldsymbol{\beta}$ and setting to zero, we have

$$\frac{\partial \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2}{\partial \boldsymbol{\beta}} = \mathbf{0}_{(p+1)\times 1} = -2\mathbf{X}^t_{(p+1)\times n}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})_{n\times 1}$$

Least Squares Estimation: Continued II

Differentiating RSS($\boldsymbol{\beta}$) with respect to $\boldsymbol{\beta}$ and setting to zero, we have

$$
\begin{aligned}
\frac{\partial \|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2}{\partial \boldsymbol{\beta}} \quad &= \quad \mathbf{0}_{(p+1)\times 1} = -2\mathbf{X}^t_{(p+1)\times n}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})_{n\times 1} \\
&\Longrightarrow \quad \mathbf{X}^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \mathbf{0} \quad \text{normal equation} \\
&\Longrightarrow \quad (\mathbf{X}^t\mathbf{X})\boldsymbol{\beta} = \mathbf{X}^t\mathbf{y} \\
&\Longrightarrow \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}
\end{aligned}
$$

Here we assume the rank of $\mathbf{X}$ is $(p+1)$ and then the inverse of the $(p+1)\times(p+1)$ matrix $(\mathbf{X}^t\mathbf{X})$ exists.

Least Squares Estimation: Continued II

Differentiating RSS($\boldsymbol{\beta}$) with respect to $\boldsymbol{\beta}$ and setting to zero, we have

$$
\begin{aligned}
\frac{\partial \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2}{\partial \boldsymbol{\beta}} &= \mathbf{0}_{(p+1)\times 1} = -2\mathbf{X}^t_{(p+1)\times n}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})_{n\times 1} \\
&\Longrightarrow \mathbf{X}^t(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0} \quad \text{normal equation} \\
&\Longrightarrow (\mathbf{X}^t\mathbf{X})\boldsymbol{\beta} = \mathbf{X}^t\mathbf{y} \\
&\Longrightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}
\end{aligned}
$$

Here we assume the rank of $\mathbf{X}$ is $(p+1)$ and then the inverse of the $(p+1)\times(p+1)$ matrix $(\mathbf{X}^t\mathbf{X})$ exists. What if rank$(X) < (p+1)$? Not a serious issue.