STAT 3690 Lecture 16

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Mar 07, 2022

What is a linear model?

- Responses are linear functions with respect to unknown parameters.

Univariate/multiple linear regression

- Interested in the relationship between random scalar Y and random q-vector $[X_1, \ldots, X_q]^\top$
- Model
  - Population version: $Y \mid X_1, \ldots, X_q \sim ([1, X_1, \ldots, X_q]\boldsymbol{\beta}, \sigma^2)$, where $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_q]^\top$, i.e.,
    * $\mathrm{E}(Y \mid X_1, \ldots, X_q) = [1, X_1, \ldots, X_q]\boldsymbol{\beta} = \beta_0 + \sum_{j=1}^q X_j\beta_j$
    * $\mathrm{var}(Y \mid X_1, \ldots, X_q) = \sigma^2$
  - Sample version $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
    * $\mathbf{Y} = [Y_1, \ldots, Y_n]^\top$ and design matrix

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{q1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nq} \end{bmatrix}_{n\times(q+1)}$$

      · Independent realizations $[Y_i, X_{i1}, \ldots, X_{iq}]^\top \sim [Y, X_1, \ldots, X_q]^\top$, $i = 1, \ldots, n$
      · $\mathrm{rk}(\mathbf{X}) = q + 1 < p + q + 1 \le n$
    * $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_n]^\top \sim (\mathbf{0}_n, \sigma^2\mathbf{I}_n)$

- Least squares (LS) estimation (no need of normality)
  - $\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$
  - $\hat{\sigma}^2_{\mathrm{LS}} = (n - q - 1)^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (n - q - 1)^{-1}\mathbf{Y}^\top(\mathbf{I} - \mathbf{H})\mathbf{Y}$
    * Hat matrix $\mathbf{H} = [h_{ij}]_{n\times n} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$
      · Symmetric
      · Idempotent: $\mathbf{H}^2 = \mathbf{H}\mathbf{H} = \mathbf{H}$
      · $\mathrm{rk}(\mathbf{H}) = \mathrm{rk}(\mathbf{X})$
      · Each eigenvalue of $\mathbf{H}$ is either zero or one

a. Prove that $\hat{\beta}_{LS} = \arg\min_{\beta} Q(\beta) = (Y - X\beta)^T (Y - X\beta) = \arg\min_{\beta} Y^TY - 2\beta^T X^T Y + \beta^T X^T X \beta$

Known: $\frac{\partial Q(\beta)}{\partial \beta} = -2X^TY + 2X^TX\beta$ (Matrix calculus)

Let $\partial Q(\beta) / \partial\beta = 0$, then $X^TY = X^TX\beta$.

So, $\hat{\beta}_{LS} = (X^TX)^{-1}X^TY$ is a stationary point, i.e., a candidate for the minimizer

Actually, for any β,

$$Q(\beta) = (Y - X\hat{\beta}_{LS} + X\hat{\beta}_{LS} - X\beta)^T (Y - X\hat{\beta}_{LS} + X\hat{\beta}_{LS} - X\beta)$$

$$= \{Y - X\hat{\beta}_{LS} + X(\hat{\beta}_{LS} - \beta)\}^T \{Y - X\hat{\beta}_{LS} + X(\hat{\beta}_{LS} - \beta)\}$$

$$= (Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS}) + \underbrace{(\hat{\beta}_{LS} - \beta)^T X^T X (\hat{\beta}_{LS} - \beta)}_{\geq 0} + 2\underbrace{(Y - X\hat{\beta}_{LS})^T X (\hat{\beta}_{LS} - \beta)}_{(1)}$$

$$(1) = (Y - HY)^T H X (\hat{\beta}_{LS} - \beta) = Y^T \underbrace{(I - H) H}_{0} X (\hat{\beta}_{LS} - \beta) = 0$$

That is, $Q(\beta) \geq (Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS}) = Q(\hat{\beta}_{LS})$ for all β.

b. Prove that $E(\hat{\sigma}^2_{LS}) = \sigma^2$

$$Y^T (I - H) Y = (Y - X\beta + X\beta)^T (I - H)(Y - X\beta + X\beta)$$

$$= (Y - X\beta)^T (I - H)(Y - X\beta) + 2\beta^T \underbrace{X^T (I - H)}_{0} (Y - X\beta) + \beta^T \underbrace{X^T (I - H)}_{0} X\beta$$

$$= tr\{(I - H)(Y - X\beta)(Y - X\beta)^T\}$$

$$\therefore E\{Y^T (I - H) Y \mid X\} = E[tr\{(I - H)(Y - X\beta)(Y - X\beta)^T\} \mid X]$$

$$= tr[(I - H) E\{(Y - X\beta)(Y - X\beta)^T \mid X\}]$$

$$= tr\{(I - H) Cov(Y - X\beta \mid X)\}$$

$$= tr\{(I - H) Cov(Y \mid X)\}$$

$$= tr\{\sigma^2 (I - H)\} = \sigma^2 (n - q - 1)$$

$\therefore E(\hat{\sigma}^2_{LS} \mid X) = \sigma^2$ for all X

$\therefore E(\hat{\sigma}^2_{LS}) = \sigma^2$

- Maximum likelihood (ML) estimation (in need of normality)
  - $\hat{\boldsymbol{\beta}}_{\rm ML} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} = \hat{\boldsymbol{\beta}}_{\rm LS}$
  - $\hat{\sigma}^2_{\rm ML} = n^{-1}\mathbf{Y}(\mathbf{I} - \mathbf{H})\mathbf{Y} = n^{-1}(n - q - 1)\hat{\sigma}^2_{\rm LS}$

$\because \ln L(X\beta, \sigma) = const - n\ln\sigma - \frac{1}{2\sigma^2}(Y - X\beta)^T (Y - X\beta)$

$\therefore$ for each σ, maximize $\ln L(X\beta, \sigma) \Leftrightarrow$ minimize $(Y - X\beta)^T (Y - X\beta)$

$\therefore \hat{\beta}_{ML} = \hat{\beta}_{LS}$

$\therefore \hat{\sigma}_{ML} = \arg\max_{\sigma} \ln L(X\hat{\beta}_{LS}, \sigma)$

Let $\partial \ln L(X\hat{\beta}_{LS}, \sigma) / \partial\sigma = -n/\sigma + \sigma^{-3}(Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS}) = 0$

Then $\hat{\sigma} = \{(Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS}) / n\}^{1/2}$ is a stationary point.

$$\because \partial^2 \ln L(X\hat{\beta}_{LS}, \sigma) / \partial\sigma^2 \Big|_{\sigma = \hat{\sigma}} = n\hat{\sigma}^{-2} - 3\hat{\sigma}^{-4}(Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS})$$

$$= \frac{n^2 - 3n^2}{(Y - X\hat{\beta}_{LS})^T (Y - X\hat{\beta}_{LS})} < 0$$

$\therefore \hat{\sigma} = \arg\max_{\sigma} \ln L(X\hat{\beta}_{LS}, \sigma) = \hat{\sigma}_{ML}$

- Inference (in need of normality)
  - $100(1-\alpha)\%$ confidence interval for $\boldsymbol{a}^\top \boldsymbol{\beta}$ with known $\boldsymbol{a} \in \mathbb{R}^{q+1}$:

$$\boldsymbol{a}^\top \hat{\boldsymbol{\beta}}_{\mathrm{ML}} \pm t_{1-\alpha/2, n-q-1} \hat{\sigma}_{\mathrm{ML}} [\boldsymbol{a}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{a}]^{1/2}$$

  - $100(1-\alpha)\%$ prediction interval for $Y_0 = \mathbf{X}_0 \boldsymbol{\beta} + \varepsilon_0$, given $\mathbf{X}_0 = [1, X_{01}, \ldots, X_{0q}] \in \mathbb{R}^{q+1}$:

$$\mathbf{X}_0 \hat{\boldsymbol{\beta}}_{\mathrm{ML}} \pm t_{1-\alpha/2, n-q-1} \hat{\sigma}_{\mathrm{ML}} [1 + \mathbf{X}_0 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}_0^\top]^{1/2}$$