DW 23/11/15

ST3451: Problem set 3

November, 2015

Problem 1 is due at class on Monday 30th November 5pm class.

1. Consider the matrix formulation of the simple linear regression model $E\{\mathbf{Y}\} = \mathbf{X}\beta$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ is an $n \times 1$ column vector of responses, $\beta = (\beta_0, \beta_1)^T$, and $\mathbf{X}$ is the $n \times 2$ design matrix having row i equal to $(1\ X_i)$. Find in terms of the X_i:

   (a) $\mathbf{X}^T\mathbf{X}$

   (b) $|\mathbf{X}^T\mathbf{X}|$

   (c) $(\mathbf{X}^T\mathbf{X})^{-1}$

   Verify that $\widehat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ gives the same least squares estimators as we derived in chapter 1 in class. What is the hat matrix in this instance? Does it have a simple form?

2. Suppose we have K regression lines

   $$Y_{ki} = \alpha_k + \beta_k X_{ki} + \epsilon_{ki} \qquad (i = 1, \ldots, n_k)$$

   where the ϵ_{ki} are independently and identically distributed as $N(0, \sigma^2)$. Write the joint model for all data

   $$\mathbf{Y} = [Y_{11}, Y_{12}, \ldots, Y_{1n_1}, \ldots, Y_{K1}, \ldots, Y_{Kn_K}]^T$$

   in the form

   $$\mathbf{Y} = \mathbf{X}\gamma + \epsilon.$$

   What are $\mathbf{X}$ and γ?

3. Consider a model where observations are time ordered $Y_1, Y_2, \ldots, Y_t, \ldots, Y_n$, where t denotes time. We have observations on p possible predictors. There is a changepoint at time $t = \tau$, so that up to time τ it is reasonable to assume that

   $$E\{Y_t\} = \mathbf{x}_t^T\beta_1, \qquad t = 1, \ldots, \tau$$

   while after time τ

   $$E\{Y_t\} = \mathbf{x}_t^T\beta_2, \qquad t = \tau + 1, \ldots, n$$

   where $\mathbf{x}_t = (1, X_{1t}, X_{2t}, \ldots, X_{pt})^T$ and the β's are $(p+1) \times 1$ parameter vectors for each of the regimes.

   (a) If we assume a constant error variance $\mathrm{var}\{\epsilon_i\} = \sigma^2$ over both regimes, find the least squares estimators of $\beta_1, \beta_2, \sigma^2$.

   (b) If we assume a different error variance $\mathrm{var}\{\epsilon_i\} = \sigma_1^2$, $i = 1, \ldots, \tau$ and $\mathrm{var}\{\epsilon_i\} = \sigma_2^2$, $i = \tau + 1, \ldots, n$ find the least squares estimators of $\beta_1, \beta_2, \sigma_1^2, \sigma_2^2$.

   (c) Suggest and approach for determining an estimate of τ if it is not known in advance.

4. If $\mathbf{H}$ is the hat matrix for a multiple linear regression model, show that $\mathbf{H}^T\mathbf{H} = \mathbf{H}$.

5. If $\mathbf{H}$ is the hat matrix for a multiple linear regression model, show that $SSE = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y}$.

6. Suppose that a multiple regression model is mis-specified, so that we have modelled the data as

$$\mathrm{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}$$

whereas in actual fact

$$\mathrm{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}$$

where $\mathbf{Z}$ is an $n \times q$ matrix of unobserved predictors and $\boldsymbol{\gamma}$ is a $q \times 1$ column vector of parameters. The least squares estimator of $\boldsymbol{\beta}$ is computed as $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{Y}$. What are the actual sampling properties of $\widehat{\boldsymbol{\beta}}$ if we assume normally distributed, uncorrelated errors? Show that the expected value of the residual error vector $\boldsymbol{\varepsilon}$ is $(\mathbf{I} - \mathbf{H})\,\mathbf{Z}\,\boldsymbol{\gamma}$.

7. (*) In the set up of question 6, suppose that the estimate

$$\widehat{\sigma}^2 = \frac{\mathbf{Y}^{\mathsf{T}}(\mathbf{I}-\mathbf{H})\mathbf{Y}}{n-p}$$

is used. Show that

$$\mathrm{E}\{\widehat{\sigma}^2\} = \sigma^2 + \frac{\boldsymbol{\gamma}^{\mathsf{T}}\mathbf{Z}^{\mathsf{T}}(\mathbf{I}-\mathbf{H})\mathbf{Z}\boldsymbol{\gamma}}{n-p} > \sigma^2$$

i.e. that the variance is over-estimated. Is it possible to design a test for $\boldsymbol{\gamma} = \mathbf{0}$?

07/02/15 ALSM 1

PROBLEM SHEET 3

Q1 $\mathbb{E}[y] = X\beta$ &R $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$

A $X^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix}$

B $|X^T X| = ad - bc = n \Sigma x_i^2 - (\Sigma x_i)^2 - nS_{xx}$

C. $(X^T X)^{-1} = \frac{1}{|X^T X|} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix}$ SWAP DIAGONALS

$= \frac{1}{nS_{xx}} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix}$

Verify that $\hat{\beta} = (X^T X)^{-1} X^T \underline{y}$ gives the same as chapter 1

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ $\hat{\beta}_1 = S_{xy}/S_{xx}$

$(X^T X)^{-1} X^T = \frac{1}{nS_{xx}} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & & x_n \end{bmatrix}$

$= \frac{1}{nS_{xx}} \begin{bmatrix} \Sigma x_i^2 - x_1 \Sigma x_i & \cdots & \Sigma x_i^2 - x_n \Sigma x_i \\ -\Sigma x_i + n x_1 & \cdots & -\Sigma x_i + n x_n \end{bmatrix}$

$\hat{\beta} = (X^T X)^{-1} X^T y$ $\frac{1}{nS_{xx}} \begin{bmatrix} \downarrow \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

$= \frac{1}{nS_{xx}} \begin{bmatrix} (\Sigma x_i^2 - x_1 \Sigma x_i) y_1 + \cdots + (\Sigma x_i^2 - x_n \Sigma x_i) y_n \\ (-\Sigma x_i + n x_1) y_1 + \cdots + (-\Sigma x_i + n x_n) y_n \end{bmatrix}$

$= \frac{1}{nS_{xx}} \begin{bmatrix} \Sigma y \Sigma x_i^2 - \Sigma x_i y_i \Sigma x_i \\ -\Sigma x_i \Sigma y_i + n \Sigma x_i y_i \end{bmatrix}$

$= \begin{bmatrix} \frac{1}{nS_{xx}}(\ ) \\ S_{xy}/S_{xx} \end{bmatrix}$

For $\hat{\beta}_0$: $\Sigma y_i \Sigma x_i^2 - \Sigma x_i y_i \Sigma x_i$

$= \Sigma y_i \Sigma x_i - (S_{xy} + \frac{\Sigma x_i \Sigma y_i}{n}) \Sigma x_i$

$= (\Sigma y_i)(\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}) - S_{xy} \Sigma x_i$

$= \Sigma y_i S_{xx} - S_{xy} \Sigma x_i$

OW 23/11/15

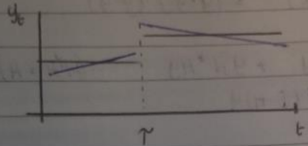
07/12/15 ALSM 1

Q3 Time ordered $y_1, \dots, y_t, y_{t+1}, \dots y_n$

Up to time τ (change point) $\mathbb{E}[y_t] = x_t^T \beta_1$ $t = 1, \dots, \tau$

$\mathbb{E}[y_t] = x_t^T \beta_2$ $t = \tau_{+1}, \dots, n$

Imagine $p = 1$



Connection to splines, in spline you constrain the lines to meet

A. Common error variance σ^2 across all times. Find $\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2$

$Q = (\hat{\beta}_1, \hat{\beta}_2)$ $\quad y_1 = \begin{bmatrix} y_1 \\ y_\tau \end{bmatrix}$ $x_1 = \begin{bmatrix} 1 & x_1 \\ 1 & x_\tau \end{bmatrix}$ Similarly y_2 and x_2

$Q: (\underline{y}_1 - x_1\hat{\beta}_1)^T(\underline{y}_1 - x_1\hat{\beta}_1) + \underbrace{(\underline{y}_2 - x_2\hat{\beta}_2)^T(\underline{y}_1 - x_2\hat{\beta}_1)}_{\text{error contribution}}$

LS Estimators: $\hat{\beta}_1 = (x_1^T x_1)^{-1} x_1^T \underline{y}_1$

$\hat{\beta}_2 = (x_2^T x_2)^{-1} x_2^T \underline{y}_2$

$\hat{\sigma}^2 = \text{pooled MSE} = \frac{Q(\hat{\beta}_1, \hat{\beta}_2)}{n-2}$ SSE

B If the variance can be different, $\hat{\sigma}_1^2, \hat{\sigma}_2^2$ would have to use a weighted LSE. Formulate a LS criterion for each regime

$Q(\hat{\beta}_1) = (\underline{y}_1 - x_1\hat{\beta}_1)^T(\underline{y}_1 - x_1\hat{\beta}_1)$

$Q(\hat{\beta}_2) = (\underline{y}_2 - x_2\hat{\beta}_2)(\underline{y}_2 - x_2\hat{\beta}_2)$

$\hat{\beta}_1$ as before $\hat{\beta}_2$ as before

$\hat{\sigma}_1^2 = \frac{Q(\hat{\beta}_1)}{\tau - 2}$ use MSE for first segment $\hat{\sigma}_2^2 = \frac{Q(\hat{\beta}_2)}{n - \tau - 2}$

C What one could do is fit the model for different values of τ (candidate) and choose $\hat{\tau}$ to minimise the Q function [or sum of Q's in second case]

Q4 $H^T H = H = X(X^T X)^{-1} X^T \Rightarrow H^T = X(X^T X)^{-1} X^T$

$H^T H = X(X^T X)^{-1} \cancel{X^T X}(X^T X)^{-1} X^T$

$= X(X^T X)^{-1} X^T = H$

1^{st} entry is $\frac{1}{nS_{xx}}\left[\Sigma(y_i)S_{xx} - S_{xy}\Sigma x_i\right]$

$= \bar{y} - S_{xy}/S_{xx}\,\bar{x} \qquad = \bar{y} - \hat{\beta}_1 \bar{x}$

LS estimates are the same

Hat Matrix $\quad H = X(X^TX)^{-1}X^T$

$$= \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \Sigma x_i^2 - x_1\Sigma x_i + & \cdots & + \Sigma x_i^2 - x_n\Sigma x_i \\ -\Sigma x_i + n x_1 + & \cdots & + -\Sigma x_i + n x_n \end{bmatrix} \frac{1}{nS_{xx}}$$

$(n \times n)$ matrix $\rightarrow$ look at the k,l entry

$$= \begin{bmatrix} 1 & x_k \end{bmatrix} \begin{bmatrix} \Sigma x_i^2 - x_l\Sigma x_i \\ -\Sigma x_i + n x_l \end{bmatrix}$$

$= \Sigma x_i^2 - x_l\Sigma x_i - x_k\Sigma x_i + n x_k x_l$

$= \Sigma x_i^2 \quad (x_l + x_k)\Sigma x_i \quad + n x_k x_l$

$= S_{xx} + \frac{(\Sigma x_i)^2}{2} - (x_l + x_k)\Sigma x_i + n x_l x_k$

$= S_{xx} + n\left[\frac{(\Sigma x_i)^2}{n} - (x_l + x_k)\frac{\Sigma x_i}{n} + x_l x_k\right]$

$= S_{xx} + n(x_l - \bar{x})(x_k - \bar{x})$

$H_{kl} = \frac{1}{n} + \frac{(x_l - \bar{x})(x_k - \bar{x})}{S_{xx}}$

Q2 k regression lines $\quad y_{ki} = a_k + \beta_k x_{ki} + \varepsilon_{ki} \quad i = 1, \ldots, n_k$

$\varepsilon_{ki} \sim N(0, \sigma^2)$ iid $\quad \underline{y} = X\underline{\gamma} + \underline{\varepsilon}$

$$\underset{\left(\sum_{i=1}^{k} n_i\right)\times 1}{\underline{y}}: \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{bmatrix} = \underset{\left(\sum_{i=1}^{k} n_i\right)\times 2k}{X}: \begin{bmatrix} 1 & x_{11} & 0 & \cdots & 0 & 0 \\ 1 & x_{12} & 0 & & & \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & x_{1n_1} & 0 & & & \\ 0 & 1 & x_{21} & & & \\ \vdots & \vdots & \vdots & & & \vdots \\ & 1 & x_{2n_2} & & & \\ & 0 & 1 & & & \\ & & 0 & & & \\ \vdots & & & & 0 & 0 \\ & & & & 1 & x_{k1} \\ 0 & 0 & 0 & \cdots & 1 & x_{kn_k} \end{bmatrix} \quad \underset{2k\times 1 \text{ matrix}}{\underline{\gamma}}: \begin{bmatrix} a_1 \\ \beta_1 \\ a_2 \\ \beta_2 \\ \vdots \\ a_k \\ \beta_k \end{bmatrix} \quad \underline{\varepsilon}: \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n_1} \\ \vdots \\ \varepsilon_{k1} \\ \vdots \\ \varepsilon_{kn_k} \end{bmatrix}$$

$\left(\sum_{i=1}^{k} n_i\right) \times 1 \qquad \left(\sum_{i=1}^{k} n_i\right) \times 2k \qquad 2k \times 1$ matrix

Q5 $H \rightarrow$ Hat matrix $\hat{y} = Hy$ $H = X(X^TX)^{-1}X^T$

SSE $= y^T(I-H)y$

SSE $= \sum \hat{\varepsilon}_i^2 = \sum (y_i - \hat{y}_i)^2 = (y-\hat{y})^T(y-\hat{y})$

$= (y-Hy)^T(y-Hy)$

$= y^Ty - y^THy - y^THy + y^TH^THy$ $(H^TH = H)$

$= y^Ty + y^THy = y^T(I-H)y$

Q6 $\mathbb{E}[y] = X\beta$ Assumed Model

$\mathbb{E}[y] = X\beta + Z\gamma$ True Model

Z - $n \times q$ matrix of "extra" predictors (would be measured already or not)

γ - $q \times 1$ vector of corresponding parameters

$\hat{\beta} = (X^TX)^{-1}X^Ty$

If we assume the "wrong" model, what are the sampling properties of $\hat{\beta}$

$\hookrightarrow \mathbb{E}[\hat{\beta}] = \beta$ $Var[\hat{\beta}] = \sigma^2(X^TX)^{-1}$

$\mathbb{E}[\hat{\beta}]$ under true model: $= \mathbb{E}[(X^TX)^{-1}X^Ty] = (X^TX)^{-1}X^T\,\mathbb{E}[y]$

$= (X^TX)^{-1}X^T(X\beta + Z\gamma)$

$= \beta + \underbrace{(X^TX)^{-1}X^TZ\gamma}_{\text{introduced a bias}}$ Make LES Biased - bias is 0 when $X=0$

$Var_T[\hat{\beta}] = Var[\overbrace{(X^TX)^{-1}X^T}^{A}y]$ $A\,Var[y]A^T$

$= (X^TX)^{-1}X^T\,Var[y]\,X(X^TX)^{-1}$

$= (X^TX)^{-1}X^T(\sigma^2 I)X(X^TX)^{-1}$

$= \sigma^2(X^TX)^{-1} \leftrightarrow$ Same as previous model

Model misspecification doesn't affect standard error estimate in this case