1.

$$CV_{(n)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - S_{ii}}\right)^2 \approx \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - \frac{1}{n}tr(\underset{\sim}{S})}\right)^2 \approx \frac{1}{n}\overbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}^{=RSS}\left(1 + \frac{2}{n}tr(\underset{\sim}{S})\right)$$

$$= \frac{1}{n}\left(RSS + 2tr(\underset{\sim}{S})\frac{1}{n}RSS\right)$$

$\because \hat{\sigma}^2 = \frac{1}{n-p-1}RSS \Rightarrow \frac{1}{n}RSS = \frac{n-p-1}{n}\hat{\sigma}^2$

$\therefore CV_{(n)} = \frac{1}{n}\left(RSS + 2tr(\underset{\sim}{S})\left(\frac{n-p-1}{n}\right)\hat{\sigma}^2\right)$

and the $C_p = \frac{1}{n}\left(RSS + 2tr(\underset{\sim}{S})\hat{\sigma}^2\right)$

2.

(a) Since the solution for β_0 is $\frac{\partial J(\beta)}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_j) = 0$

$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}\beta_j x_{ij}) = \bar{y} - \sum_{j=1}^{p}\beta_j \bar{x}_j \Rightarrow \bar{y} - \bar{\underset{\sim}{X}}\beta$

The problem reduces to

$$\| y - \underset{\sim}{X}\beta - \bar{y} + \bar{\underset{\sim}{X}}\beta\|^2 + \lambda\|\beta\|^2 = \|(y - \bar{y}) - (\underset{\sim}{X} - \bar{\underset{\sim}{X}})\beta\|^2 + \lambda\|\beta\|^2$$

$$= \| y_c - \underset{\sim}{X}_c\beta\|^2 + \lambda\|\beta\|^2$$

The ridge regression solution is obtained by minimizing

$RSS(\lambda) = (y_c - \underset{\sim}{X}_c\beta)^T(y_c - \underset{\sim}{X}_c\beta) + \lambda\beta^T\beta$

$\frac{\partial RSS}{\partial \beta} = 0 \Rightarrow -2\underset{\sim}{X}_c^T y_c + 2\underset{\sim}{X}_c^T\underset{\sim}{X}_c\beta + 2\lambda\beta = 0$

$$\Rightarrow \hat{\beta}_\lambda = \left(\underset{\sim}{X}_c^T\underset{\sim}{X}_c + \lambda \mathbb{I}\right)^{-1}\underset{\sim}{X}_c^T y_c$$

(b)

$\hat{\beta}_\lambda = (VDU^TUDV^T + \lambda VV^T)^{-1}VDU^T y_c$

$= V(DD + \lambda \mathbb{I})^{-1}DU^T y_c$

$= \sum_{j=1}^{p}\frac{d_j}{d_j^2 + \lambda}v_j u_j^T y_c \quad \Rightarrow \|\hat{\beta}_\lambda\||_{\lambda>0} \leq \|\hat{\beta}_\lambda\||_{\lambda=0}$

(c)

$$\overline{err} = \frac{1}{n}\|y_c - \hat{y}_c\|^2 = \frac{1}{n}\|y_c - X_c^T\hat{\beta}_\lambda\|^2$$

$$= \frac{1}{n}\|y - X_c^T(X_c^TX_c + \lambda I)^{-1}X_c^T y_c\|^2$$

$$= \frac{1}{n}\|(I - X_c^T(X_c^TX_c + \lambda I)^{-1}X_c^T)y_c\|^2$$

$$= \frac{1}{n} y_c^T(I - X_c^T(X_c^TX_c + \lambda I)^{-1}X_c^T)^T(I - X_c^T(X_c^TX_c + \lambda I_c)^{-1}X_c^T)y_c$$

By taking SVD of X_c to yield $X_c = UDV^T = \sum_{j=1}^{p} d_j u_j v_j^T$

$$X_c(X_c^TX_c + \lambda I)^{-1}X_c^T$$

$$= UDV^T(VDU^TUDV^T + \lambda VV^T)^{-1}VDU^T$$

$$= UD(DD + \lambda I)^{-1}DU^T = \sum_{j=1}^{p}\frac{d_j^2}{d_j^2+\lambda} u_j u_j^T$$

Then, $I - X_c(X_c^TX_c + \lambda I)^{-1}X_c^T = UU^T - \sum_{j=1}^{p}\frac{d_j^2}{d_j^2+\lambda}u_j u_j^T = \sum_{j=1}^{p}\frac{\lambda}{d_j^2+\lambda}u_j u_j^T$

Since $\sum_{j=1}^{p}\frac{\lambda}{d_j^2+\lambda}u_j u_j^T$ increases as λ increases

Training error is an increasing function of λ

Ch6. Model. selection.

3 $\quad C_p : \frac{1}{n} \cdot (RSS + 2d\hat{\sigma}^2)$

(a) $\quad \hat{\sigma} = \frac{1}{n-p-1} \cdot RSS$

① $p = 2$. chose all features.

$$\begin{array}{cccc} & \beta_0 & \beta_1 & \beta_2 \end{array}$$

$\beta = (X^T X)^{-1} X^T Y = [\ 0\ ,\ -0.4\ ,\ 0.2\ ]$ $\rightarrow$ $y = [\ 1\ , -1\ ,\ 0\ ]$, $\hat{y} = [\ 1\ , -1\ ,\ 0\ ]$

$RSS = 0$. $\longrightarrow$ $\hat{\sigma} = 0$ # Best Model.

② $p = 1$, chose first feature. x_{i1}

$\beta = [\ 0\ ,\ -0.5\ ]$ $\longrightarrow$ $y = [\ 1\ , -1\ ,\ 0\ ]$, $\hat{y} = [\ 1\ , -0.5\ ,\ 0.5\ ]$

$RSS = 0.501 \longrightarrow C_p = \frac{1}{3} \cdot (0.501 + 2 \cdot 1 \cdot 0)$

③ $p = 1$, chose second feature x_{i2}

$\beta = [\ 0\ ,\ 0.286\ ]$ $\longrightarrow$ $y = [\ 1\ , -1\ ,\ 0\ ]$, $\hat{y} = [0.286, -0.857, 0.571]$

$RSS = 0.858 \longrightarrow C_p = \frac{1}{3} \cdot (0.858 + 2 \cdot 1 \cdot 0)$

ref. Ch6. P16.

(b) $(-2, 1, 1)$ $(1, -3, -1)$ $(1, 2, 0)$

$$\min_{\beta_1,\beta_2} \frac{1}{2}\|y - X^T\beta\|^2 + \lambda\|\beta\|_1$$

$$= \min_{\beta_1,\beta_2} \frac{1}{2}y^Ty - y^TX\beta + \frac{1}{2}\beta^T\beta + \lambda\|\beta\|_1$$

無關β.

$$= \min_{\beta_1,\beta_2} -y^TX\beta + \frac{1}{2}\|\beta\|^2 + \lambda\|\beta\|_1$$

let $y^TX = \alpha \rightarrow \min_{\beta_1,\beta_2} -\alpha\beta + \frac{1}{2}\|\beta\|^2 + \lambda\|\beta\|_1$

$\alpha = [-3, 4]$. $\rightarrow \min_{\beta_1,\beta_2} L(\beta)$

$$\rightarrow \frac{\partial L(\beta)}{\partial \beta_i} = 0$$

$\rightarrow$ case 1 : $\beta_i > 0$, $i = 1, 2$ $\quad \beta_i = \alpha - \lambda \Rightarrow \beta = [-13, -6]$

$\rightarrow$ case 2 : $\beta_i < 0$, $i = 1, 2$ $\quad \beta_i = \alpha + \lambda \Rightarrow \beta = [7, 14]$.

$\rightarrow$ not satisfied case 1, case 2. $\rightarrow \beta_1 = \beta_2 = 0$

$\beta_0 = 0$ #

(Recall that $\beta_0 = \bar{y} - \sum_{j=1}^{p} \beta_j \bar{x}_j$)

$(\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i)$ $(\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij})$

4.

(a)
$$\begin{matrix}(-1, & -1, & -1)\\(0, & 0, & 0)\\(2, & 1, & -1)\end{matrix} \Rightarrow \begin{cases}\bar{x}_{i1} = \frac{1}{3}\\ \bar{x}_{i2} = 0\\ \bar{x}_{i3} = -\frac{2}{3}\end{cases}$$

Let

$$\underline{X}_1 = \begin{bmatrix}-1\\0\\2\end{bmatrix} - \bar{x}_{i1} = \begin{bmatrix}-\frac{4}{3}\\-\frac{1}{3}\\\frac{5}{3}\end{bmatrix};\ \underline{X}_2 = \begin{bmatrix}-1\\0\\1\end{bmatrix} - 0 = \begin{bmatrix}-1\\0\\1\end{bmatrix};\ \underline{X}_3 = \begin{bmatrix}-1\\0\\-1\end{bmatrix} - \left(-\frac{2}{3}\right) = \begin{bmatrix}-\frac{1}{3}\\\frac{2}{3}\\-\frac{1}{3}\end{bmatrix}$$

$$X = [\underline{X}_1\ \underline{X}_2\ \underline{X}_3] = \begin{bmatrix}-\frac{4}{3} & -1 & -\frac{1}{3}\\ -\frac{1}{3} & 0 & \frac{2}{3}\\ \frac{5}{3} & 1 & -\frac{1}{3}\end{bmatrix}$$

$$\phi_1 = \underset{\|\phi_1'\|^2=1}{\arg\max}\ \mathrm{Var}(\phi_{11}' X_1 + \phi_{21}' X_2 + \phi_{31}' X_3),\quad \phi_1 = [\phi_{11}\ \phi_{21}\ \phi_{31}]^T$$

$= $ eigenvector corresponding to max. eigenvalue of $X^T X$

$$\tilde{X} = [\tilde{\underline{X}}_1\ \tilde{\underline{X}}_2\ \tilde{\underline{X}}_3]^T = X - \phi_1 \phi_1^T X$$

$$\tilde{\phi}_2 = \underset{\|\phi_2'\|^2=1}{\arg\max}\ \mathrm{Var}(\tilde{\phi}_{21}' \tilde{X}_1 + \tilde{\phi}_{22}' \tilde{X}_2 + \tilde{\phi}_{23}' \tilde{X}_3)$$

$= $ eigenvector corresponding to the 2nd largest eigenvalue of $X^T X$

$$X^T X = \begin{bmatrix}\frac{14}{3} & 3 & -\frac{1}{3}\\ 3 & 2 & 0\\ -\frac{1}{3} & 0 & \frac{2}{3}\end{bmatrix}:$$

eigenvalue	eigenvector
0	$[-0.53\ \ 0.8\ \ -0.27]^T$
0.704	$[-0.11\ \ 0.25\ \ 0.96]^T$
6.63	$[-0.84\ \ -0.54\ \ 0.05]^T$

$\therefore\ W_{i1} = -0.84 x_{i1} - 0.54 x_{i2} + 0.05 x_{i3}$: 1st P.C.

$W_{i2} = -0.11 x_{i1} + 0.25 x_{i2} + 0.96 x_{i3}$: 2nd P.C.

(b)

1st P.C. : $W_{i1} = -0.84 x_{i1} - 0.54 x_{i2} + 0.05 x_{i3}$ — 2nd P.C: $W_{i2} = -0.11 x_{i1} + 0.25 x_{i2} + 0.96 x_{i3}$

$$\begin{cases}(-1,-1,-1) : W_{11} = 1.43\\ (0,0,0) : W_{21} = 0\\ (2,1,-1) : W_{31} = -2.27\end{cases}\qquad \begin{cases}(-1,-1,-1) : W_{12} = -1.1\\ (0,0,0) : W_{22} = 0\\ (2,1,-1) : W_{32} = -0.93\end{cases}$$

$\bar{W} = -0.28$, $\sigma^2 = 2.325$ $\qquad$ $\bar{w} = -0.6767$, $\sigma^2 = 0.2338$

3rd P.C.: $W_{i3} = -0.53\, X_{i1} + 0.8\, X_{i2} - 0.27\, X_{i3}$

$$\begin{cases} (-1,-1,-1) : W_{13} \approx 0 \\ (0,0,0) : W_{23} = 0 \\ (2,1,-1) : W_{33} = 0.01 \end{cases} \qquad \bar{W} = 0.0033 \qquad \sigma^2 = 2.22\times 10^{-5}$$

(c)

$$\phi_{j1}^{PLS} = \arg\max_{\phi_{j1}} E[(Y - \phi_{j1} X_j)^2]$$

$$\Rightarrow \phi_{j1}^{PLS} = \frac{\underline{x}_j^T \underline{y}}{\underline{x}_j^T \underline{x}_j}, \quad \underline{y} \triangleq [y_1\ y_2\ y_3]^T = [-3\ 1\ 5]^T$$

$$\therefore \phi_{11} = \frac{\underline{x}_1^T \underline{y}}{\underline{x}_1^T \underline{x}_1} = 2.5714$$

$\phi_{21} = 4$

$\phi_{31} = 0$

$$\begin{cases} \text{1st P.C. direction}, \ \phi_1 = [-0.84\ \ -0.54\ \ 0.05]^T \\ \text{1st PLS direction}\ \ \phi_1^{PLS} = [2.5714\ \ 4\ \ 0]^T \end{cases}$$

5. For $X < \xi_1 \Rightarrow (X-\xi_1)_+ = (X-\xi_2)_+ = (X-\xi_3)_+ = 0$

$$\Rightarrow f(X) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$$

$\because$ Linear boundary condition $(f(X) = AX + B)$

$\therefore \beta_2 = \beta_3 = 0 \quad \cdots ①$

For $X > \xi_K \Rightarrow f(X) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \sum_{k=1}^{K} \beta_{k+3} (X^3 - 3X^2 \xi_k + 3X\xi_k^2 - \xi_k^3)$

$\because$ Linear boundary condition $(f(X) = AX + B)$

$$\therefore \begin{cases} \beta_2 + \sum_{k=1}^{K} \beta_{k+3} \cdot (-3\xi_k) = 0 \\ \beta_3 + \sum_{k=1}^{K} \beta_{k+3} \cdot 1 = 0 \end{cases} \underset{\text{by}①}{\Rightarrow} \begin{cases} \sum_{k=1}^{K} \beta_{k+3} \xi_k = 0 \\ \sum_{k=1}^{K} \beta_{k+3} = 0 \end{cases} \#$$

6. (a) For $x \le \xi$

$$\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 = a_1 + b_1 x + c_1 x^2 + d_1 x^3$$

$\Rightarrow a_1 = \beta_0,\ b_1 = \beta_1,\ c_1 = \beta_2,\ d_1 = \beta_3$ #

(b) For $x > \xi$

$$\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x-\xi)^3$$

$$= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x^3 - 3x^2\xi + 3x\xi^2 - \xi^3)$$

$$= (\beta_0 - \beta_4 \xi^3) + (\beta_1 + 3\beta_4 \xi^2) x + (\beta_2 - 3\beta_4 \xi) x^2 + (\beta_3 + \beta_4) x^3$$

$$= a_2 + b_2 x + c_2 x^2 + d_2 x^3$$

$\Rightarrow a_2 = \beta_0 - \beta_4 \xi^3,\ b_2 = \beta_1 + 3\beta_4 \xi^2,\ c_2 = \beta_2 - 3\beta_4 \xi,\ d_2 = \beta_3 + \beta_4$ #

(c) $f_1(\xi) = \beta_0 + \beta_1 \xi + \beta_2 \xi^2 + \beta_3 \xi^3$

$$f_2(\xi) = (\beta_0 - \beta_4 \xi^3) + (\beta_1 + 3\beta_4 \xi^2)\xi + (\beta_2 - 3\beta_4 \xi)\xi^2 + (\beta_3 + \beta_4)\xi^3$$

$$= \beta_0 \cancel{- \beta_4 \xi^3} + \beta_1 \xi \cancel{+ 3\beta_4 \xi^3} + \beta_2 \xi^2 - 3\beta_4 \xi^3 + \beta_3 \xi^3 \cancel{+ \beta_4 \xi^3}$$

$$= \beta_0 + \beta_1 \xi + \beta_2 \xi^2 + \beta_3 \xi^3$$

$$= f_1(\xi) \quad *$$

(d) $f_1'(\xi) = \beta_1 + 2\beta_2 \xi + 3\beta_3 \xi^2$

$$f_2'(\xi) = (\beta_1 + 3\beta_4 \xi^2) + 2(\beta_2 - 3\beta_4 \xi)\xi + 3(\beta_3 + \beta_4)\xi^2$$

$$= \beta_1 + 3\beta_4 \xi^2 + 2\beta_2 \xi - 6\beta_4 \xi^2 + 3\beta_3 \xi^2 \cancel{+ 3\beta_4 \xi^2}$$

$$= \beta_1 + 2\beta_2 \xi + 3\beta_3 \xi^2$$

$$= f_1'(\xi) \quad *$$

(e) $f_1''(\xi) = 2\beta_2 + 6\beta_3 \xi$

$$f_2''(\xi) = 2(\beta_2 - 3\beta_4 \xi) + 6(\beta_3 + \beta_4)\xi$$

$$= 2\beta_2 \cancel{- 6\beta_4 \xi} + 6\beta_3 \xi + 6\beta_4 \xi$$

$$= 2\beta_2 + 6\beta_3 \xi$$

$$= f_1''(\xi)$$

Therefore, $f(x)$ is indeed a cubic spline,