High-dimensional data analysis, Midterm exam #2: **[+28 points]**

Name 林[illegible]

+25

- Not only answer but also calculation

+6

1. **[+6]** Consider a linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$,

$$\text{where } \mathbf{y} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}.$$

+1

(1) [+1] Obtain the centered design matrix $\mathbf{X}^C$ such that the sum of each column of $\mathbf{X}^C$ is 0.

$\bar{X}_1 = \frac{1+0+0}{3} = \frac{1}{3}$

$\bar{X}_2 = \frac{0+1+0}{3} = \frac{1}{3}$

$\bar{X}_3 = \frac{1+1+0}{3} = \frac{2}{3}$

$$\Rightarrow X^c = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{2}{3} \end{bmatrix} \#$$

+1

(2) [+1] $(\mathbf{X}^C)^T\mathbf{X}^C =$

$$(X^c)^T X^c = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \#$$

+4

(3) [+4] Obtain the ridge estimator $\hat{\boldsymbol{\beta}}^{ridge} = \arg\min\{(\mathbf{y} - \mathbf{X}^C\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}^C\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}^T\boldsymbol{\beta}\}$ for $\lambda = 1/3$.

$$\hat{\boldsymbol{\beta}}^{ridge} = \begin{bmatrix} \hat{\beta}_1^{ridge} \\ \hat{\beta}_2^{ridge} \\ \hat{\beta}_3^{ridge} \end{bmatrix} = (X^{cT}X^c + \lambda I_3)^{-1} X^{cT} y = \begin{bmatrix} \frac{3}{2} & \frac{3}{4} & -\frac{3}{4} \\ \frac{3}{4} & \frac{3}{2} & -\frac{3}{4} \\ -\frac{3}{4} & -\frac{3}{4} & \frac{3}{2} \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{4} \\ 0 \\ -\frac{3}{4} \end{bmatrix} \#$$

Derivations

$$\frac{\partial}{\partial \beta}\left[(y - X^c\beta)^T(y - X^c\beta) + \lambda\beta^T\beta\right]$$

$$= -2X^{cT}(y - X^c\beta) + 2\lambda\beta \overset{\text{let}}{=} 0.$$

$$\Rightarrow \tilde{\beta}^{ridge} = (X^{cT}X^c + \lambda I_3)^{-1} X^{cT} y \ \#$$

$$\Rightarrow (X^{cT}X^c + \lambda I_3)^{-1} = \begin{bmatrix} 1 & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & 1 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{3}{2} & \frac{3}{4} & -\frac{3}{4} \\ \frac{3}{4} & \frac{3}{2} & -\frac{3}{4} \\ -\frac{3}{4} & -\frac{3}{4} & \frac{3}{2} \end{bmatrix}$$

and $\frac{\partial^2}{\partial \beta^2}\left[(y - X^c\beta)^T(y - X^c\beta) + \lambda\beta^T\beta\right]$

$= X^{cT}X^c + \lambda I_3 \geq 0$

$$\Rightarrow X^{cT} y = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}$$

$\because \forall a \in \mathbb{R}^3, a \neq 0$

$a^T(X^{cT}X^c + \lambda I_3)a = a^T X^{cT}X^c a + \lambda a^T a \geq \lambda a^T a \geq 0.$

47 [+8] Let $\mathbf{y}|\boldsymbol{\beta} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I)$ and $\boldsymbol{\beta} \sim N(0, \tau^2 I)$.

+3 1) [+4] Derive the density of $\boldsymbol{\beta}|\mathbf{y}$.

$$f(\beta|y) = \frac{f(y|\beta)\cdot f(\beta)}{f(y)} \propto f(y|\beta)\cdot f(\beta)$$

$$\propto \exp\left\{\frac{-1}{2}(y-X\beta)^T\frac{I}{\sigma^2}(y-X\beta)\right\}\exp\left\{\frac{-1}{2}\beta^T\frac{I}{\tau^2}\beta\right\}.$$

$$\propto \exp\left\{\frac{-1}{2\sigma^2}\left[-2y^TX\beta + \beta^TX^TX\beta + \frac{\sigma^2}{\tau^2}\beta^T\beta\right]\right\}.$$

$$\propto \exp\left\{\frac{-1}{2\sigma^2}\left[\beta^T(X^TX+\frac{\sigma^2}{\tau^2}I)\beta - 2y^TX(X^TX+\frac{\sigma^2}{\tau^2}I)^{-1}(X^TX+\frac{\sigma^2}{\tau^2}I)\beta\right\}.$$

$$\propto \exp\left\{\frac{-1}{2}\left[\beta-(X^TX+\frac{\sigma^2}{\tau^2}I)^{-1}X^Ty\right]^T\frac{(X^TX+\lambda I)}{\sigma^2}\left[\beta-(X^TX+\frac{\sigma^2}{\tau^2}I)^{-1}X^Ty\right]\right\}$$ $f(\beta|x)=?$

$\Rightarrow$

$\beta|y \sim N\left((X^TX+\lambda I)^{-1}X^Ty,\ \sigma^2(X^TX+\lambda I)^{-1}\right)$ //

where $\lambda = \frac{\sigma^2}{\tau^2}$

+1 2) [+1] Show that $\hat{\boldsymbol{\beta}}^{ridge} = (X^TX+\lambda I)^{-1}X^T\mathbf{y}$ is equal to the posterior mean $E[\boldsymbol{\beta}|\mathbf{y}]$.

by (1)

$E[\beta|y] = (X^TX+\lambda I)^{-1}X^Ty = \hat{\beta}^{ridge}$ //

+3 3) [+3] Derive the generalized ridge estimator $E[\boldsymbol{\beta}|\mathbf{y}]$ under $\mathbf{y}|\boldsymbol{\beta} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I)$ and $\boldsymbol{\beta} \sim N(0, W^{-1})$, where W is any symmetric matrix.

$$f(\beta|y) \propto f(y|\beta)\cdot f(\beta)$$

$$\propto \exp\left\{\frac{-1}{2\sigma^2}(y-X\beta)^T(y-X\beta) - \frac{1}{2}\beta^TW\beta\right\}.$$

$$\propto \exp\left\{\frac{-1}{2\sigma^2}\left[-2y^TX\beta + \beta^TX^TX\beta + \sigma^2\beta^TW\beta\right]\right\}.$$

$$\propto \exp\left\{\frac{-1}{2\sigma^2}\left[\beta^T(X^TX+\sigma^2W)\beta - 2y^TX(X^TX+\sigma^2W)^{-1}(X^TX+\sigma^2W)\beta\right]\right\}.$$

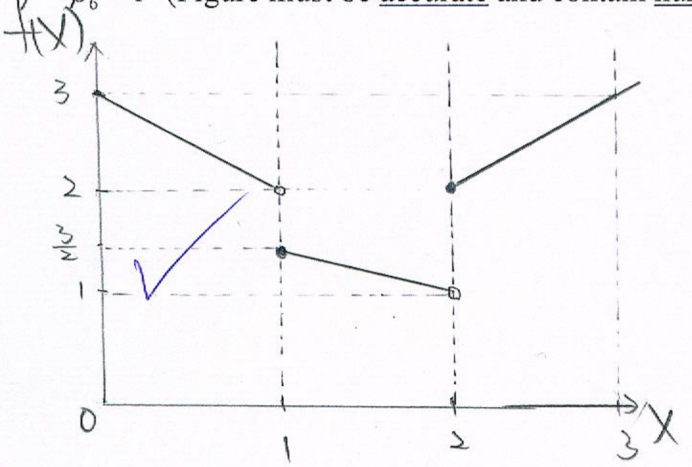
$$\propto \exp\left\{\frac{-1}{2}\left[y-(X^TX+\sigma^2W)^{-1}X^Ty\right]^T\frac{(X^TX+\sigma^2W)}{\sigma^2}\left[y-(X^TX+\sigma^2W)^{-1}X^Ty\right]\right\}.$$

$\Rightarrow \beta|y \sim N\left((X^TX+\sigma^2W)^{-1}X^Ty,\ \sigma^2(X^TX+\sigma^2W)^{-1}\right).$

$\Rightarrow E[\beta|y] = (X^TX+\sigma^2W)^{-1}X^Ty$ //

+5 **Q3 [+6]** Consider a piecewise linear bases expansion $f(X)=\sum_{m=1}^{6}\beta_m h_m(X)$, where $h_1(X)=\mathbf{I}(X<\xi_1)$, $h_2(X)=\mathbf{I}(\xi_1\le X<\xi_2)$, $h_3(X)=\mathbf{I}(\xi_2\le X)$, $h_{m+3}(X)=Xh_m(X)$, m=1, 2, 3. Let $\xi_1=1$ and $\xi_2=2$.

+2 **(1) [+2]** Draw the figure of $f(X)$ when $\beta_1=3$, $\beta_2=2$, $\beta_3=0$, $\beta_4=-1$, $\beta_5=-0.5$, and $\beta_6=1$ (Figure must be <u>accurate</u> and contain <u>numerical values of all points</u>)



$X<1$: $f(X)=\beta_1+\beta_4X=3-X$

$\Rightarrow f(0)=3,\ f(1)=2$

$1\le X<2$: $f(X)=\beta_2+\beta_5X=2-\frac{1}{2}X$

$\Rightarrow f(1)=\frac{3}{2},\ f(2)=1$

$X\ge 2 \Rightarrow f(X)=\beta_3+\beta_6X=X$

$\Rightarrow f(2)=2$

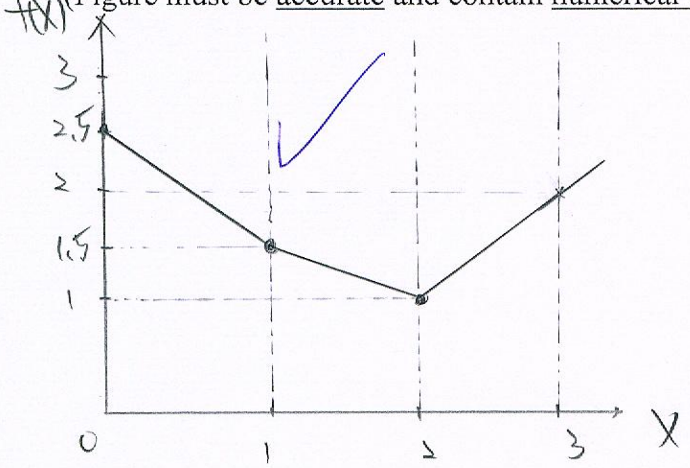
+1 **(2) [+1]** Discuss some disadvantage of using $f(X)$ in the above figure.

It is not continuous.

~~If~~ the data is continuous, piecewise linear expansion may not be useful.

+2 **(3) [+2]** We impose constraints that $f(X)$ is continuous at $\xi_1=1$ and $\xi_2=2$. Draw the figure of $f(X)$ when $\beta_1=2.5$, $\beta_4=-1$, $\beta_5=-0.5$, and $\beta_6=1$.

(Figure must be <u>accurate</u> and contain <u>numerical values of all points</u>)



$X<1$: $f(X)=2.5-X$

$\Rightarrow f(0)=2.5,\ f(1)=1.5$

$1\le X<2$: $f(X)=2-\frac{1}{2}X$:

$f(1)=1.5,\ f(2)=1$

$X\ge 2$: $f(X)=-1+X$ $(\beta_3=-1)$

$f(2)=1$

+0 **(4) [+1]** Discuss some disadvantage of using $f(X)$ in the above figure.

? Although f(X) is continuous, It may not fit the data well.

why ? Explain more.

+7 **Q4[+8]** Consider a model: $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$, where $\sum_{i=1}^{n} x_i = 0$

+1 (1) [+1] Write down the LSE $\begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \end{bmatrix}$.

$$X^TX = \begin{bmatrix} n & 0 \\ 0 & \sum x_i^2 \end{bmatrix} \rightarrow (X^TX)^{-1} = \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum x_i^2} \end{bmatrix}, \quad X^Ty = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \Rightarrow \begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum x_i^2} \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} = \begin{bmatrix} \bar y \\ \frac{\sum x_i y_i}{\sum x_i^2} \end{bmatrix} \ \cancel{\#}$$

+1 (2) [+2] Define t-statistics for testing $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$.

$$\because E(\hat\beta_1) = \sum_{i=1} x_i y_i, \quad Var(\hat\beta_1) = \sigma^2 \frac{1}{\sum_{i=1} x_i^2} = \frac{\sigma^2}{\sum_{i=1} x_i^2} \quad (\text{replace } \sigma^2 \text{ to } \hat\sigma^2).$$

$$\Rightarrow t = \frac{\hat\beta_1 - \beta_1^{\nearrow 0}}{\hat\sigma\sqrt{1/\sum x_i^2}} = \frac{\hat\beta_1}{\hat\sigma\sqrt{1/\sum x_i^2}} \quad \text{with d.f.} = n-2 \ \cancel{\#}$$

$\hat\sigma^2 = ?$ Define

(-1)

+2 (3) [+2] Show that F- statistics and t-statistics for $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ are equivalent.

(Proof must be clear and detailed)

model 0 $\Rightarrow y = \beta_0 + \varepsilon$

model 1 $\Rightarrow y = \beta_0 + \beta_1 X + \varepsilon$, $\sum x_i = 0$.

$$\Rightarrow RSS_0(\beta_0) = \sum_{i=1}^{n} (y_i - \bar y)^2$$

$$RSS_1(\beta_0, \beta_1) = \sum_{i=1}^{n} [y_i - (\bar y + \hat\beta_1 x_i)]^2 = \sum_{i=1}^{n} (y_i - \bar y)^2 - \hat\beta_1^2 \sum x_i^2$$

$$\Rightarrow F = \frac{[RSS_0(\beta_0) - RSS_1(\beta_0,\beta_1)]/1}{RSS_1(\beta_0,\beta_1)/n-2} = \frac{\hat\beta_1^2 \sum_{i=1} x_i^2}{\hat\sigma^2} = \left(\frac{\hat\beta_1}{\hat\sigma\sqrt{1/\sum x_i^2}}\right)^2 = t^2 \ \cancel{\#}$$

+3 (4) [+3] Draw the 95% confidence set for $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ (Figure must be detailed)

$$C_{\beta_0,\beta_1} = \{ (\beta - \hat\beta)^T (X^TX)(\beta - \hat\beta) \le \chi^2(0.95) \cdot \hat\sigma^2 \} \quad \because \beta \sim N(\beta, \sigma^2 (X^TX)^{-1})$$

$$\Rightarrow [\beta_0 - \hat\beta_0, \beta_1 - \hat\beta_1] \begin{bmatrix} n & 0 \\ 0 & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 - \hat\beta_0 \\ \beta_1 - \hat\beta_1 \end{bmatrix} \le \chi^2(0.95)\hat\sigma^2 \}. \quad \text{where} \begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \end{bmatrix} = \begin{bmatrix} \bar y \\ \frac{\sum x_i y_i}{\sum_{i=1} x_i^2} \end{bmatrix}$$

$$\Rightarrow n(\beta_0 - \hat\beta_0)^2 + \sum_{i=1}^{n} x_i^2 (\beta_1 - \hat\beta_1)^2 \le \chi^2(0.95) \cdot \hat\sigma^2 \}$$