32

Final exam, High-dimensional Data Analysis, 2018 Spring [+32 points]

Not only answer but also calculation **Name:** ______

+8 **Q1 Ridge regression [+8]:** Consider a model <u>without an intercept</u>:

$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$, $i = 1,\dots,n$. Assume $\sum_{i=1}^{n} x_{i1} = \sum_{i=1}^{n} x_{i2} = 0$ and $\sum_{i=1}^{n} x_{i1}^2 = \sum_{i=1}^{n} x_{i2}^2 = 1$.

Let $r_{12} = \sum_{i=1}^{n} x_{i1}x_{i2}$, $r_{1y} = \sum_{i=1}^{n} x_{i1}Y_i$, and $r_{2y} = \sum_{i=1}^{n} x_{i2}Y_i$, and assume $r_{12} = r_{1y} = r_{2y} = 1/2$.

+3 **(1) [+3]** Derive the <u>ridge</u> estimators $\hat\beta_{1\lambda}$ and $\hat\beta_{2\lambda}$ (your answer should be simplified)

$L = (y - X\beta)^T(y - X\beta) - \lambda\beta^T\beta$

$L' = -2X(y - X\beta) - 2\lambda\beta \overset{set}{=} 0$

$\Rightarrow \hat\beta = (X^TX + \lambda I)^{-1}X^Ty$

$X = \begin{bmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix}$ $X^T = \begin{bmatrix} x_{11} & \cdots & x_{n1} \\ x_{12} & \cdots & x_{n2} \end{bmatrix}$ $X^TX = \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} \\ \sum_{i=1}^n x_{i1}x_{i2} & \sum_{i=1}^n x_{i2}^2 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$

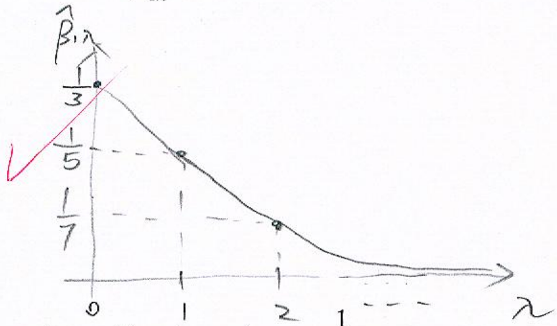
$\hat\beta = \left(\begin{bmatrix} 1 & \frac12 \\ \frac12 & 1 \end{bmatrix} + \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)^{-1}\begin{bmatrix} \frac12 \\ \frac12 \end{bmatrix}$ $X^Ty = \begin{bmatrix} \sum_{i=1}^n x_{i1}y_i \\ \sum_{i=1}^n x_{i2}y_i \end{bmatrix} = \begin{bmatrix} \frac12 \\ \frac12 \end{bmatrix}$

$= \begin{bmatrix} 1+\lambda & \frac12 \\ \frac12 & 1+\lambda \end{bmatrix}^{-1}\begin{bmatrix} \frac12 \\ \frac12 \end{bmatrix} = \begin{bmatrix} \frac{1}{2\lambda+3} \\ \frac{1}{2\lambda+3} \end{bmatrix}$

$\therefore \hat\beta_{1\lambda} = \hat\beta_{2\lambda} = \frac{1}{2\lambda+3}$

+2 **(2) [+2]** Draw the ridge trace for $\hat\beta_{1\lambda}$

λ	0	1	2	$\cdots$	∞
$\hat\beta_{1\lambda}$	$\frac13$	$\frac15$	$\frac17$	$\cdots$	0



+2 **(3) [+2]** Choose λ by setting $df_\lambda = 1$ and $r_{12} = \frac{1}{2}$.

$df_\lambda = trace[X(X^TX + \lambda I)^{-1}X^T]$

$= trace[(X^TX + \lambda I)^{-1}X^TX]$

$= trace\left(\frac{1}{(2\lambda+3)(2\lambda+1)}\begin{bmatrix} 4\lambda+3 & 2\lambda \\ 2\lambda & 4\lambda+3 \end{bmatrix}\right) = \frac{2(4\lambda+3)}{(2\lambda+3)(2\lambda+1)} \overset{set}{=} 1 \Rightarrow \lambda^2 = \frac34$

$\Rightarrow \lambda = \frac{\sqrt3}{2}$

+1 **(4) [+1]** Calculate the ridge estimates under the chosen value of λ.

When $\lambda = \frac{\sqrt3}{2}$

$\hat\beta_{1\lambda} = \hat\beta_{2\lambda} = \frac{1}{2\times\frac{\sqrt3}{2}+3} = \frac{3-\sqrt3}{6}$

+8

Q2 Splines [+8]: Let $f(x)=\sum_{j=1}^{K}\beta_j M_j(x)$, where $M_j(x)$'s are the spline basis functions. For a knot sequence $\xi_1<\xi_2<\xi_3$ with $\Delta=\xi_2-\xi_1=\xi_3-\xi_2$, let $z_i(t)=(t-\xi_i)/\Delta$ for $i=1$, 2, and 3.

+2 **(1) [+2]** Write $\int f''(x)^2dx$ in term of a quadratic form for $\boldsymbol{\beta}=(\beta_1,\dots,\beta_K)^T$.

$f(x)=\sum_{j=1}^{K}\beta_j M_j(x)$

$f''(x)=\sum_{j=1}^{K}\beta_j M_j''(x)$

$$\int f''(x)^2dx = \int \sum_{j=1}^{K}\beta_j M_j(x)\sum_{j'=1}^{K}\beta_{j'}M_{j'}(x)dx = \sum_{j=1}^{K}\sum_{j'=1}^{K}\beta_j\beta_{j'}\int M_j(x)M_{j'}(x)dx$$

$$= \int \sum_{j=1}^{K}\sum_{j'=1}^{K}\beta_j\beta_{j'}M_j(x)M_{j'}(x)dx = \beta\int M_j''(x)M_{j'}''(x)dx\,\beta^T$$

+6 **(2) [+6]** Calculate $\int M_4''(t)^2\,dt$,

where $M_4(t)=\dfrac{\mathbf{I}(\xi_1\le t<\xi_2)}{2\Delta}z_1(t)^3+\dfrac{\mathbf{I}(\xi_2\le t<\xi_3)}{2\Delta}\{-7z_2(t)^3+3z_2(t)^2+3z_2(t)+1\}$

$I_1=I(\xi_1\le t<\xi_2)$

$I_2=I(\xi_2\le t<\xi_3)$

$$M_4'(t)=\frac{I_1}{2\Delta^2}3z_1(t)^2+\frac{I_2}{2\Delta^2}\{-21z_2(t)^2+6z_2(t)+3\}$$

$$M_4''(t)=\frac{3I_1}{\Delta^3}z_1(t)+\frac{I_2}{\Delta^3}\cdot\{-21z_2(t)+3\}$$

$$\int M_4''(t)^2dt=\int\left[\frac{3I_1}{\Delta^3}z_1(t)+\frac{I_2}{\Delta^3}\{-21z_2(t)+3\}\right]^2dt$$

$$=\int\frac{9I_1}{\Delta^6}z_1(t)^2dt+\int\frac{I_2}{\Delta^6}\{-21z_2(t)+3\}^2dt$$

$$=\frac{9}{\Delta^6}\int_{\xi_1}^{\xi_2}z_1(t)^2dt+\frac{9}{\Delta^6}\int_{\xi_2}^{\xi_3}\{7z_2(t)-1\}^2dt$$

$49z_2^2-14z_2+1$

$$=\frac{9}{\Delta^6}\left\{\left[\frac{\Delta}{3}\left(\frac{t-\xi_1}{\Delta}\right)^3\right]_{\xi_1}^{\xi_2}+\left[\frac{49\Delta}{3}\left(\frac{t-\xi_2}{\Delta}\right)^3-\frac{14\Delta}{2}\left(\frac{t-\xi_2}{\Delta}\right)^2+\Delta\left(\frac{t-\xi_2}{\Delta}\right)\right]_{\xi_2}^{\xi_3}\right\}$$

$$=\frac{9}{\Delta^5}\left\{\frac{1}{3}+\frac{49}{3}-\frac{14}{2}+1\right\}=\frac{96}{\Delta^5}$$

+8

Q3 Compound covariates [+8]: Consider a model <u>without an intercept</u>:

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2), \; i = 1,\ldots,n.$$

Assume $\sum_{i=1}^{n} x_{i1} = \sum_{i=1}^{n} x_{i2} = 0$ and $\sum_{i=1}^{n} x_{i1}^2 = \sum_{i=1}^{n} x_{i2}^2 = n$. Let $\mathrm{CC}_i = \hat\beta_1 x_{i1} + \hat\beta_2 x_{i2}$ be a compound covariate.

+2 **(1) [+2]** Derive $\hat\beta_1$ and $\hat\beta_2$ in the CC.

$\hat\beta_1 = \underset{\beta_1}{\mathrm{argmax}} \sum_{i=1}^{n} (y_i - \beta_1 x_{i1})^2$

$L = \sum (y_i - \beta_1 x_{i1})^2$

$L' = -2 \sum x_{i1}(y_i - \beta_1 x_{i1}) \overset{set}{=} 0 \Rightarrow \hat\beta_1 = \frac{\sum x_{i1} y_i}{\sum x_{i1}^2} = \frac{\sum x_{i1} y_i}{n}$

the same as $\hat\beta_1$

$\hat\beta_2 = \frac{\sum x_{i2} y_i}{\sum x_{i2}^2} = \frac{\sum x_{i2} y_i}{n}$

+1 **(2) [+1]** Calculate $\sum_{i=1}^{n} \mathrm{CC}_i$.

$\sum_{i=1}^{n} cc_i = \sum_{i=1}^{n} (\hat\beta_1 x_{i1} + \hat\beta_2 x_{i2}) = \hat\beta_1 \sum x_{i1} + \hat\beta_2 x_{i2} = 0$

+5 **(3) [+5]** Let $\hat Y_i^{CC} = \hat\gamma \mathrm{CC}_i$ be a predictor of Y_i. Derive $\hat\gamma$ in terms of $\hat\beta_1$, $\hat\beta_2$, and $r_{12} = \frac{1}{n}\sum_{i=1}^{n} x_{i1} x_{i2}$.

$L = \sum_{i=1}^{n} (y_i - \gamma cc_i)^2$

$L' = -2 \sum_{i=1}^{n} cc_i (y_i - \gamma cc_i) \overset{set}{=} 0$

$$\Rightarrow \hat\gamma = \frac{\sum_{i=1}^{n} cc_i y_i}{\sum_{i=1}^{n} cc_i^2}$$

$$= \frac{\sum_{i=1}^{n} (\hat\beta_1 x_{i1} + \hat\beta_2 x_{i2}) y_i}{\sum_{i=1}^{n} (\hat\beta_1 x_{i1} + \hat\beta_2 x_{i2})^2}$$

$$= \frac{\hat\beta_1 \sum_{i=1}^{n} (x_{i1} y_i) + \hat\beta_2 \sum_{i=1}^{n} (x_{i2} y_i)}{\hat\beta_1^2 \sum_{i=1}^{n} x_{i1}^2 + \hat\beta_2^2 \sum_{i=1}^{n} x_{i2}^2 + 2\hat\beta_1\hat\beta_2 \sum x_{i1} x_{i2}}$$

$\sum_{i=1}^{n} (x_{i1} y_i) = \hat\beta_1 \sum x_{i1}^2 = n\hat\beta_1$

$\sum_{i=1}^{n} (x_{i2} y_i) = \hat\beta_2 \sum x_{i2}^2 = n\hat\beta_2$

$$= \frac{n\hat\beta_1^2 + n\hat\beta_2^2}{n\hat\beta_1^2 + n\hat\beta_2^2 + 2n r_{12} \hat\beta_1 \hat\beta_2}$$

$$= \frac{\hat\beta_1^2 + \hat\beta_2^2}{\hat\beta_1^2 + \hat\beta_2^2 + 2\hat\beta_1\hat\beta_2 r_{12}}$$

+8

Q4 Cross-validation [+8]: Data is given below:

y_i	x_i	h_i	$\hat{y}_i$
0	2	$\frac{25}{72}$	$\frac{13}{12}$
2	-2	$\frac{25}{72}$	$\frac{29}{12}$
3	0	$\frac{1}{8}$	$\frac{7}{4}$
4	-1	$\frac{13}{72}$	$\frac{25}{12}$
2	0	$\frac{1}{8}$	$\frac{7}{4}$
1	-2	$\frac{25}{72}$	$\frac{29}{12}$
0	1	$\frac{13}{72}$	$\frac{17}{12}$
2	2	$\frac{25}{72}$	$\frac{13}{12}$

$$X = \begin{bmatrix} 1 & 2 \\ 1 & -2 \\ 1 & 0 \\ 1 & -1 \\ 1 & 0 \\ 1 & -2 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad X^TX = \begin{bmatrix} 8 & 0 \\ 0 & 18 \end{bmatrix} \qquad (X^TX)^{-1} = \begin{bmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{18} \end{bmatrix}$$

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & -2 & 0 & -1 & 0 & -2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 3 \\ 4 \\ 2 \\ 1 \\ 0 \\ 2 \end{bmatrix} \qquad X^Ty = \begin{pmatrix} 14 \\ -6 \end{pmatrix}$$

+2 **(1) [+2]** Compute the leverage h_i in the table above. Write the fraction (e.g., 1/2, not 0.5)

$$h_i = x_i^T (X^TX)^{-1} x_i$$

① : $(1\ 2)\begin{pmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{18} \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{25}{72}$

② : $(1\ -2)\begin{pmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{18} \end{pmatrix}\begin{pmatrix} 1 \\ -2 \end{pmatrix} = \frac{25}{72}$

③ : $(1\ 0)\begin{pmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{18} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{8}$

④ : $(1\ 1)\begin{pmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{18} \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{13}{72}$

⑤ $(1\ -1)\begin{pmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{18} \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{7}$

+2 **(2) [+2]** Compute the predictor $\hat{y}_i$ in the above table

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

$$= \begin{bmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{18} \end{bmatrix}\begin{pmatrix} 14 \\ -6 \end{pmatrix} = \begin{pmatrix} \frac{7}{4} \\ -\frac{1}{3} \end{pmatrix} \Rightarrow \hat{y}_i = \frac{7}{4} - \frac{1}{3}x_i$$

+2 **(3) [+2]** Compute the training MSE

$$MSE = \frac{1}{8}\sum_{i=1}^{8}(y_i - \hat{y}_i)^2$$

$$= \frac{1}{8}\Big\{(0-\tfrac{13}{12})^2 + (2-\tfrac{29}{12})^2 + (3-\tfrac{7}{4})^2 + (4-\tfrac{25}{12})^2 + (2-\tfrac{7}{4})^2 + (1-\tfrac{29}{12})^2 + (0-\tfrac{17}{12})^2 + (2-$$

$$= \frac{1}{8}\Big\{(-\tfrac{13}{12})^2 + (-\tfrac{5}{12})^2 + (\tfrac{5}{4})^2 + (\tfrac{23}{12})^2 + (\tfrac{1}{4})^2 + (-\tfrac{17}{12})^2 + (-\tfrac{17}{12})^2 + (\tfrac{11}{12})^2\Big\}$$

$$= \frac{23}{16}$$

+2 **(4) [+2]** Compute the testing MSE by LOOCV (use h_i)

$$LOOCV = \frac{1}{8}\sum_{i=1}^{8}\left(\frac{y_i - \hat{y}_i}{1-h_i}\right)^2$$

$$= \frac{1}{8}\Big\{\frac{169}{64} + \frac{25}{64} + \frac{100}{49} + \frac{19044}{3481} + \frac{4}{49} + \frac{10164}{2209} + \frac{10404}{3481} + \frac{4356}{2209}\Big\}$$

$$= 2.5369$$