Name: Vo Ngoc Bich Uyen

HOME ASSIGNMENT 3

Points for problems: 1, 1, 1, 0.5, 0.5
Total: 4

Problem 1: Let us recall that, for any matrix $A \in \mathbb{R}^{m\times n}$ with SVD $A = U^T \Sigma V = \sum_{j=1}^{rank(A)} \sigma_j u_j v_j^T$, its pseudo-inverse A^+ is defined as $A^+ = \sum_{j=1}^{rank(A)} \frac{1}{\sigma_j} v_j u_j^T$.

Let $A^+ = \lim_{\lambda\to+0} (A^TA + \lambda I_n)^{-1} A^T$. Show that $A^+ = A^+$

$\longrightarrow$ Proof:

+) A $m\times n$ matrix $A = U^T\Sigma V \Rightarrow V_{n\times n}$ is an unitary matrix; i.e. $I_n = V.V^T = \sum_{j=1}^{n} v_j v_j^T$

+) We have

$$A^+ = \lim_{\lambda\to+0} (A^TA + \lambda I_n^{-1})^{-1} A^T = \lim_{\lambda\to+0}\left\{\left[\sum_{j=1}^{rank(A)}(\sigma_j^2 v_j u_j^T . u_j . v_j^T + \lambda v_j v_j^T)\right]^{-1} . \left(\sum_{j=1}^{rank(A)} \sigma_j v_j u_j^T\right)\right\}$$

$$= \lim_{\lambda\to+0}\left\{\left[\sum_{j=1}^{rank(A)} (\sigma_j^2+\lambda) v_j v_j^T\right]^{-1} . \left(\sum_{j=1}^{rank(A)} \sigma_j . v_j u_j^T\right)\right\} = \lim_{\lambda\to+0}\left(\sum_{j=1}^{rank(A)} \frac{\sigma_j}{\sigma_j^2+\lambda} . v_j . v_j^T . v_j . u_j^T\right)$$

$$= \sum_{j=1}^{rank(A)} \frac{1}{\sigma_j} v_j . u_j^T = A^+ \qquad \square$$

Problem 2: For any $\lambda > 0$, define the ridge regression estimate as

$$\hat{\beta}_\lambda^{ridge} \in \underset{\beta}{\operatorname{argmin}} \frac{1}{n}\|Y - X\beta\|^2 + \lambda\|\beta\|^2$$

Prove that $\hat{\beta}^{ls} = (X^TX)^+X^TY = \lim_{\lambda\to+0} \hat{\beta}_\lambda^{ridge}$

$\longrightarrow$ Proof:

Denote that $C(\beta) = \frac{1}{n}\|Y - X\beta\|^2 + \lambda\|\beta\|^2$

Firstly, we have to prove that $C(\beta)$ is convex.

Indeed, we have

$$C(\lambda\beta + (1-\lambda)\alpha) = \frac{1}{n}\|\lambda Y + (1-\lambda)Y - X[\lambda\beta + (1-\lambda)\alpha]\|^2 + \lambda\|\lambda\beta + (1-\lambda)\alpha\|^2$$

$$= \frac{\lambda^2}{n}\|Y - X\beta\|^2 + \frac{(1-\lambda)^2}{n}.\|Y - X\alpha\|^2 + \frac{2\lambda(1-\lambda)}{n}(Y - X\beta)^T(Y - X\alpha) +$$

$$+ \lambda^3\|\beta\|^2 + \lambda(1-\lambda)^2\|\alpha\|^2 + 2\lambda^2(1-\lambda)\|\beta.\alpha\| \le \left[a^Tb \le \frac{\|a\|^2}{2} + \frac{\|b\|^2}{2}\right] \le$$

$$\le \frac{\lambda^2}{n}\|Y - X\beta\|^2 + \frac{(1-\lambda)^2}{n}\|Y - X\alpha\|^2 + \frac{\lambda(1-\lambda)}{n}(\|Y - X\beta\|^2 + \|Y - X\alpha\|^2) + \lambda^3\|\beta\|^2 +$$

$$+ \lambda(1-\lambda)^2\|\alpha\|^2 + \lambda^2(1-\lambda)(\|\beta\|^2 + \|\alpha\|^2) = \frac{\lambda}{n}\|Y - X\beta\|^2 + \lambda^2\|\beta\|^2 + \frac{(1-\lambda)}{n}\|Y - X\alpha\|^2 + \lambda(1-\lambda)\|\alpha\|^2$$

$$= \lambda C(\beta) + (1-\alpha)C(\alpha)$$

+) Base on the property of convex function: $\hat{\beta}_\lambda^{ridge} \in \underset{\beta}{\operatorname{argmin}}\, C(\beta) \iff \nabla C(\hat{\beta}) = 0$

$\Rightarrow \nabla C(\hat{\beta}) = \frac{2}{n} X^T(X\hat{\beta} - Y) + 2\lambda\hat{\beta} = 0 \qquad (\hat{\beta} = \hat{\beta}_\lambda^{ridge})$

$\iff \quad (X^TX + n\lambda)\hat{\beta} = X^TY$

$\Rightarrow \quad \hat{\beta} = (X^TX + n\lambda I)^+ X^TY$

$\Rightarrow \quad \lim_{\lambda\to+0} \hat{\beta}_\lambda^{ridge} = \lim_{\lambda\to+0} \hat{\beta} = \lim_{\lambda\to+0} [(X^TX + n\lambda I)^+ X^TY] = (X^TX)^+X^TY = \hat{\beta}^{ls}$ □

<u>Problem 3:</u> We say that a design matrix $X \in \mathbb{R}^{n\times d}$ satisfies the restricted isometry property with constants $s \in \mathbb{N}$ and $\delta \in (0,1)$, if

$$(1-\delta)\|\beta\|^2 \le \frac{1}{n}\|X\beta\|^2 \le (1+\delta)\|\beta\|^2 \text{ for all } \beta \in \mathbb{R}^d \text{ such that } \|\beta\|_0 \le s$$

Assume that X is ε-incoherent. Show that, for any $s \in \{1, 2, \dots, [1/\varepsilon]\}$, X satisfies the restricted isometry property with constants s and εs

$\rightarrow$ <u>Proof:</u>

+) The design matrix X is ε-incoherent, so $\left\|\frac{1}{n}X^TX - I_d\right\|_\infty \le \varepsilon \qquad (1)$

+) For any $s \in \{1, 2, \dots, [1/\varepsilon]\}$; $\varepsilon s \in \{\varepsilon, 2\varepsilon, \dots, 1\} \in (0,1]$, we have to prove that

$$(1-\varepsilon s)\|\beta\|^2 \le \frac{1}{n}\|X\beta\|^2 \le (1+\varepsilon s)\|\beta\|^2 \text{ for all } \beta \in \mathbb{R}^d \text{ such that } \|\beta\|_0 = \sum_{j=1\,(\beta_j\neq 0)}^{d} \mathbb{1} \le s$$

+) Indeed, let consider $\frac{\|X\beta\|^2}{n\|\beta\|^2}$

$$\frac{\|X\beta\|^2}{n\|\beta\|^2} = \frac{\|\beta\|^2 + \beta^T\left(\frac{1}{n}X^TX - I_d\right)\beta}{\|\beta\|^2} = 1 + \frac{\beta^T\left(\frac{1}{n}X^TX - I_d\right)\beta}{\|\beta\|^2}$$

- We have $\left|\beta^T\left(\frac{1}{n}X^TX - I_d\right)\beta\right| \le \sum_{i,j=1}^{d} |\beta_i||\beta_j| \underbrace{\left\|\frac{1}{n}X^TX - I_d\right\|_\infty}_{\le \varepsilon \text{ (from (1))}} \le \varepsilon . \|\beta\|_1^2 \qquad (2)$

- And $\|\beta\|_1 = \sum_{j=1}^{d}|\beta_j| \underset{\text{Cauchy-Schwarz inequality}}{\le} \sqrt{\sum_{j=1\,(\beta_j\neq 0)}^{d} \mathbb{1}} . \sqrt{\sum_{j=1}^{d}\beta_j^2} \le \sqrt{s}.\|\beta\| \qquad (3)$

- From (2) and (3): $\left|\beta^T\left(\frac{1}{n}X^TX - I_d\right)\beta\right| \le \varepsilon s.\|\beta\|^2 \Rightarrow -\varepsilon s.\|\beta\|^2 \le \beta^T\left(\frac{1}{n}X^TX - I_d\right)\beta \le \varepsilon s\|\beta\|^2$

Hence, $\quad 1 - \varepsilon s \le \frac{\|X\beta\|^2}{n\|\beta\|^2} \le 1 + \varepsilon s$

or $\quad (1-\varepsilon s)\|\beta\|^2 \le \frac{1}{n}\|X\beta\|^2 \le (1+\varepsilon s)\|\beta\|^2$ for all $\beta \in \mathbb{R}^d$; $\|\beta\|_0 \le s$ □

Problem 4: Let $\hat{\beta}^{bic} \in \underset{\beta}{\text{argmin}} \left\{ \frac{1}{n} \|Y - X\beta\|^2 + \lambda^2 \|\beta\|_0 \right\}$

and assume that $\frac{X^T X}{n} = I_d$. Show that the Bayesian information criterion admits a closed from representation in this case: $\hat{\beta}_j^{bic} = \frac{1}{n} X_j^T Y \, \mathbb{1}(|X_j^T Y| > n\lambda)$; $j \in \{1, \dots, d\}$ where $X_1, \dots, X_d$ are the columns of X.

$\rightarrow$ Proof:

Consider $\underset{\beta}{\text{argmin}} \left\{ \overbrace{\frac{1}{n} \|Y - X\beta\|^2 + \lambda^2 \|\beta\|_0}^{f} \right\}$, we can see that both $\frac{1}{n}\|Y - X\beta\|^2$ and $\lambda^2 \|\beta\|_0$ are element wise. Thus, the problem can be solved per element.

1) $\frac{1}{n} \|Y - X\hat{\beta}\|^2 = 0$

$$\Rightarrow \begin{cases} \hat{\beta}_j^{bic} = \frac{1}{n} X_j^T Y \text{ ; indeed, } \frac{1}{n} \left\| Y - \frac{X.X^T Y}{n} \cancel{\beta} \right\|^2 = \|Y - Y\| = 0 \\ f(\hat{\beta}_j^{bic}) = \lambda^2 \|\beta\|_0 \geq \lambda^2 \end{cases} \quad (1)$$

Not true

$\Rightarrow$ The loss of function is λ^2 when $\hat{\beta}_j^{bic} = \frac{1}{n} X_j^T Y$

2) $\lambda^2 \|\beta\|_0 = \lambda^2 \sum_{j=1}^{d} \mathbb{1}(\beta_j \neq 0) = 0$

$$\Rightarrow \begin{cases} \hat{\beta}_j^{bic} = 0 \\ f(\hat{\beta}_j^{bic}) = \frac{1}{n} \|Y\|^2 = \frac{1}{n^2} X^T X . \|Y\|^2 = \frac{\|X^T Y\|^2}{n^2} \end{cases} \quad (2)$$

$\Rightarrow$ The loss of function is $\|X^T Y\|^2$ when $\hat{\beta}_j^{bic} = 0$

Because of the difference in loss from both functions (1 & 2), the higher loss is chosen:

$$\hat{\beta}_j^{bic} = \begin{cases} \frac{1}{n} X_j^T Y & \text{if } \|X^T Y\|^2 > n^2 \lambda^2 \text{ or } |X^T Y| > n\lambda \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow \hat{\beta}_j^{bic} = \frac{1}{n} X_j^T Y . \mathbb{1}(|X_j^T Y| > n\lambda) \quad \square$$

Problem 5: Let a design matrix $X \in \mathbb{R}^{n \times d}$ be such that its restricted eigenvalue satisfies the inequality $\underset{\substack{J \subseteq \{1, \dots, d\} \\ \# |J| \leq s}}{\min} RE(0, J) = \lambda > 0$

Prove or disprove that X satisfies the restricted isometry property.

$\rightarrow$ Proof:

+) Let's consider matrix $\Psi_n = \frac{1}{n} X^T X$, we have

$$\min_{\substack{\beta \in \mathbb{R}^d \\ \beta \neq 0}} \left(\frac{\beta^T \Psi_n \beta}{\|\beta\|^2} \right) = \min_{\substack{\beta \in \mathbb{R}^d \\ \beta \neq 0}} \left(\frac{\|X\beta\|^2}{n\|\beta\|^2} \right) > 0$$

+) For $1 \leq s \leq d$, $\Phi_{min}(s) = \min_{\substack{x \in \mathbb{R}^d \\ 1 \leq \mu(x) \leq s}} \left(\frac{x^T \Psi_n x}{\|x\|^2} \right) = \min_{\substack{\beta \in \mathbb{R}^d \\ \beta \neq 0}} \left(\frac{\|X\beta\|^2}{n\|\beta\|^2} \right)$

+) For $1 \leq s_1, s_2 \leq d$, $\Theta_{s_1,s_2} = \max \left\{ \frac{c_1^T X_{I_1}^T X_{I_2} c_2}{n\|c_1\|^2 \|c_2\|^2} ; I_1 \cap I_2 = \emptyset, |I_i| \leq m_i, c_i \in \mathbb{R}^{I_i} \setminus \{0\}, i = \{1,2\} \right\}$

+) $RE(c, J) = \inf \left\{ \frac{\|X\beta\|^2}{n\|\beta_J\|^2} : \beta \in \mathbb{R}^d, \underbrace{\|\beta_{J^c}\|_1 \leq c\|\beta_J\|_1}_{(*)} \right\}$

+) $\frac{1}{n} \|X\beta\|^2 = \frac{1}{n} \|X\beta_J + X\beta_{J^c}\|^2 \geq \frac{1}{n} \|X\beta_J\|^2 + \frac{2}{n} \beta_J^T X^T X \beta_{J^c}$

$$\geq \Phi_{min}(s). \|\beta_J\|^2 - 2\Theta_{1,1} \|\beta_J\|_1 \|\beta_{J^c}\|_1$$

[From (*)] $\geq \Phi_{min}(s). \|\beta_J\|^2 - 2\,\Theta_{1,1}.c. \|\beta_J\|_1^2$

$[\|\beta_J\|_1 \leq \sqrt{s} \cdot \|\beta_J\|$ (Cauchy-Schwarz)$]$ $\geq \Phi_{min}(s) \|\beta_J\|^2 - 2\,\Theta_{1,1}.c.s. \|\beta_J\|^2$

$$\geq \left(\Phi_{min}(s) - 2\,\Theta_{1,1}.c.s \right). \|\beta_J\|^2$$

$$\Rightarrow \frac{\|X\beta\|^2}{n\|\beta_J\|^2} \geq \Phi_{min}(s) - 2.\Theta_{1,1}.c.s > 0$$

where $\Theta_{1,1} < \frac{1}{(1+2c).s}$

Hence, $\min_{\substack{J \subseteq \{1,\dots,d\} \\ |J| \leq s}} RE(0, J) = \Phi_{min}(s) - 0 = \min_{\substack{\beta \in \mathbb{R}^d \\ \beta \neq 0}} \left(\frac{\|X\beta\|^2}{n\|\beta\|^2} \right) = \lambda > 0$

$\Rightarrow \frac{\|X\beta\|^2}{n\|\beta\|^2} \geq \lambda$ $\rightarrow$ Can not conclude that X satisfies the RIP based on this inequality.

I do not see why the restricted isometry property
is not satisfied in your counterexample.