Zhenming Wang

Net ID: zw1806

Problem 1.

①: $E[\hat{w}] = (X^T X)^{-1} \cdot X^T \cdot E(y).$

$= (X^T X)^{-1} \cdot X^T \cdot E(Xw + \varepsilon)$

$= (X^T X)^{-1} \cdot (X^T \cdot X) E(w) +$

$= I \cdot w. = w.$

②: $Var(\hat{w}) = Var((X^T X)^{-1} X^T y)$

$= Var((X^T X)^{-1} \cdot X^T \cdot (Xw + \varepsilon))$

$= Var(w + (X^T X)^{-1} \cdot X^T \cdot \varepsilon)$

$= Var((X^T X)^{-1} \cdot X^T \cdot \varepsilon)$

$= Var(X^{-1} \cdot (X^T)^{-1} \cdot X^T \cdot \varepsilon)$

$= Var(X^{-1} \cdot \varepsilon)$

$= X^{-1} \cdot Var(\varepsilon) \cdot (X^{-1})^T$

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

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

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

Problem 2.

①: $J(\theta) = \|X\theta - y\|_2^2 + \lambda\|\theta\|_1$

$= \|x_1\theta_1 + x_2\theta_2 + X_r\theta_r - y\|_2^2 + \lambda|\theta_1| + \lambda|\theta_2| + \lambda\|\theta_r\|_1$

since x_1 and x_2 are repeated features.

$\rightarrow J(\theta) = \|x_1(\theta_1+\theta_2) + X_r\theta_r - y\|_2^2 + \lambda|\theta_1| + \lambda|\theta_2| + \lambda\|\theta_r\|_1$

Now, $\hat{\theta} = \begin{pmatrix} a \\ b \\ r \end{pmatrix}$ is a minimizer of $J(\theta)$.

To prove a and b have the same sign, let's assume for the sake of contradictions that a and b do not have the same sign.

Thus, $\lambda|a| + \lambda|b|$

$|a| + |b| > |a+b|$

$\rightarrow J(\hat{\theta}) = \|x_1(a+b) + X_r \cdot r - y\|_2^2 + \lambda|a| + \lambda|b| + \lambda\|r\|_1$

$< \|x_1(a+b) + X_r \cdot r - y\|_2^2 + \lambda|a+b| + \lambda\|r\|_1$

this contradicts the fact that $\hat{\theta} = (a, b, r^T)^T$ is a minimizer of $J(\theta)$. Hence, a and b must be equal sign or one of them is 0.

Next, since $(c, d, r^T)^T$ is also a minimizer of $J(\theta)$

Thus, $\|x_1(a+b) + X_r \cdot r - y\|_2^2 + \lambda|a| + \lambda|b| + \lambda\|r\|_1$

$= \|x_1(a+b) + X_r \cdot r - y\|_2^2 + \lambda|a+b| + \lambda\|r\|_1$ --- ①

$① = \|x_1(c+d) + X_r \cdot r - y\|_2^2 + \lambda|c| + \lambda|d| + \lambda\|r\|_1.$

According to previous proof, c and d must also have the same sign / one of them is zero,

Thus, $\|x_1(a+b) + X_r \cdot r - y\|_2^2 + \lambda|a+b| + \lambda\|r\|_1$

$= \|x_1(c+d) + X_r \cdot r - y\|_2^2 + \lambda|c+d| + \lambda\|r\|_1.$

Therefore, $a+b = c+d$.

②: Denote Objective function of Ridge Regression as $J(\theta)$.

$$J(\theta) = \|X\theta - y\|_2^2 + \lambda\|\theta\|_2^2.$$

$$= \|x_1\theta_1 + x_2\theta_2 + X_r \cdot \theta_r - y\|_2^2 + \lambda \cdot \theta_1^2 + \lambda\theta_2^2 + \lambda\|\theta_r\|^2.$$

$\hat{\theta} = \begin{pmatrix} a \\ b \\ r \end{pmatrix}$ minimizes $J(\theta)$

$$\rightarrow J(\hat{\theta}) = \|x_1 a + x_2 b + X_r \cdot r - y\|_2^2 + \lambda a^2 + \lambda b^2 + \lambda\|r\|^2$$

$$= \|x_1(a+b) + X_r \cdot r - y\|_2^2 + \lambda(a^2+b^2) + \lambda\|r\|^2$$

Denote $a+b=S$.

$J(\hat{\theta}) = \|X_1 \cdot S + X_r \cdot r - y\|_2^2 + \lambda (S^2 - 2a(S-a) + r^2)$

It is minimized when $a = \frac{S}{2}$,

Hence, for $\hat{\theta}$, $a=b$.