Exercise 1

(a) We want to find the $\hat{\theta}$ that minimizes

$$L(\theta) = \sum_{n=1}^{N} (y_n - \theta^T x_n)^2 + \lambda \|\theta\|^2$$

For this purpose, we will equate the derivative $\frac{\partial L(\theta)}{\partial \theta}$ to 0.

$$\frac{\partial L(\theta)}{\partial \theta} = 0 \Rightarrow 2\sum_{n=1}^{N} (y_n - \hat{\theta}^T x_n)(-x_n) + 2\lambda\hat{\theta} = 0 \Rightarrow$$

$$\Rightarrow -\sum_{n=1}^{N} y_n x_n + \sum_{n=1}^{N} (\hat{\theta}^T x_n) x_n + \lambda\hat{\theta} = 0 \Rightarrow$$

$$\Rightarrow \sum_{n=1}^{N} (x_n x_n^T)\hat{\theta} + \lambda I\hat{\theta} = \sum_{n=1}^{N} y_n x_n \Rightarrow$$

$$\Rightarrow \left(\sum_{n=1}^{N} (x_n x_n^T) + \lambda I\right)\hat{\theta} = \sum_{n=1}^{N} y_n x_n \quad (A)$$

(b) We have shown that (Homework 1):

- $\sum_{n=1}^{N} x_n x_n^T = X^T X$, where $X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix}$ (*)
- $\sum_{n=1}^{N} y_n x_n = X^T y$, where $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$ and X as above

Thus, $(A) \Rightarrow (X^T X + \lambda I)\hat{\theta} = X^T y \Rightarrow \hat{\theta} = (X^T X + \lambda I)^{-1} X^T y$

(*) $\Rightarrow X^T X = [x_1 \; x_2 \; \dots \; x_N] \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix} = \sum_{n=1}^{N} x_n x_n^T$

$X^T y = [x_1 \; x_2 \; \dots \; x_N] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \sum_{n=1}^{N} y_n x_n = \sum_{n=1}^{N} x_n y_n$

Exercise 2

(a) Since $\hat{\theta}_{MVU}$ is an unbiased estimator of θ_0, we know that

$$E[\hat{\theta}_{MVU}] = \theta_0$$

Therefore, $MSE_{\hat{\theta}_{MVU}} = E[(\hat{\theta} - E(\hat{\theta}))^2] + (E(\hat{\theta}) - \theta_0)^2 \Rightarrow$

$$\Rightarrow MSE_{\hat{\theta}_{MVU}} = E[(\hat{\theta}_{MVU} - E(\hat{\theta})_{MVU})^2]$$

Bias is equal to zero and MSE is equal to the variance of the estimator $\hat{\theta}_{MVU}$.

(b) $\theta_b = (1+a)\,\theta_{MVU} \Rightarrow E[\theta_b] = E[(1+a)\theta_{MVU}] \Rightarrow$

$$\Rightarrow E[\theta_b] = (1+a)\,E[\theta_{MVU}] \xrightarrow{E[\theta_{MVU}]=\theta_0} E[\theta_b] = (1+a)\,\theta_0$$

Since $a \neq 0 \Rightarrow E[\theta_b] \neq \theta_0$, and therefore θ_b is a biased estimator of θ_0.

(c) $MSE_{\hat{\theta}_{MVU}} = E[(\hat{\theta}_{MVU} - E(\hat{\theta})_{MVU})^2]$ (question (a))

In order for $MSE_{\hat{\theta}_{MVU}}$ to be 0, the variance of $\hat{\theta}_{MVU}$ needs to be 0.

This cannot be achieved for finite N (data sets), because every data set will have an estimator that will be different from the others.

(d) $\theta_b = (1+a)\,\theta_{MVU}$

$$MSE_{\theta_b} = E[(\theta_b - E(\theta_b))^2] + (E(\theta_b) - \theta_0)^2 \Rightarrow$$

$$MSE_{\theta_b} = E[((1+a)\,\theta_{MVU} - E[(1+a)\theta_{MVU}])^2] + (E[(1+a)\theta_{MVU}] - \theta_0)^2 =$$

$$= E[((1+a)\theta_{MVU} - (1+a)E[\theta_{MVU}])^2] + (E[\theta_{MVU} + a\theta_{MVU}] - \theta_0)^2 =$$

$$= E[(1+a)^2(\theta_{MVU} - E[\theta_{MVU}])^2] + (E[\theta_{MVU}] + aE[\theta_{MVU}] - \theta_0)^2$$

$$= (1+a)^2 E[(\theta_{MVU} - E[\theta_{MVU}])^2] + (aE[\theta_{MVU}])^2$$

$$= (1+a)^2 MSE_{\theta_{MVU}} + (a\theta_0)^2 \Rightarrow$$

$$\Rightarrow MSE_{\theta_b} = (1+a)^2 MSE_{\theta_{MVU}} + (a\theta_0)^2$$

(e) We search the values of a for which:

$$MSE_{\theta_b} < MSE_{\theta_{MVU}} \Rightarrow$$

$$\Rightarrow (1+a)^2 MSE_{\theta_{MVU}} + (a(\theta_0))^2 < MSE_{\theta_{MVU}} \Rightarrow$$

$$\Rightarrow MSE_{\theta_{MVU}} + 2a MSE_{\theta_{MVU}} + a^2 MSE_{\theta_{MVU}} - MSE_{\theta_{MVU}} + a^2\theta_0^2 < 0 \Rightarrow$$

$$\Rightarrow a^2(MSE_{\theta_{MVU}} + \theta_0^2) + 2a MSE_{\theta_{MVU}} < 0$$

The above is equal to 0 for $a = 0$ or $a = \dfrac{-2MSE_{\theta_{MVU}}}{MSE_{\theta_{MVU}} + \theta_0^2}$

Therefore it is less than 0 for $a \in \left(-2\dfrac{MSE_{\theta_{MVU}}}{MSE_{\theta_{MVU}} + \theta_0^2}, 0\right)$

(f) $\dfrac{-2MSE_{\theta_{MVU}}}{MSE_{\theta_{MVU}} + \theta_0^2} < a < 0 \Rightarrow 1 - \dfrac{2MSE_{\theta_{MVU}}}{MSE_{\theta_{MVU}} + \theta_0^2} < a+1 < 1 \Rightarrow$

$\Rightarrow |a+1| < 1$ and since $\theta_b = (1+a)\theta_{MVU}$ we have that

$|\theta_b| < |\theta_{MVU}|$

(g) $MSE_{\theta_b} = (1+a)^2 MSE_{\theta_{MVU}} + a^2\theta_0^2$

$$\frac{\partial MSE_{\theta_b}}{\partial a} = 0 \Rightarrow 2(1+a) MSE_{\theta_{MVU}} + 2a\theta_0^2 = 0 \Rightarrow$$

$$MSE_{\theta_{MVU}} + a MSE_{\theta_{MVU}} + a\theta_0^2 = 0 \Rightarrow$$

$$MSE_{\theta_{MVU}} + a(MSE_{\theta_{MVU}} + \theta_0^2) = 0 \Rightarrow$$

$$\Rightarrow a^* = \frac{-MSE_{\theta_{MVU}}}{MSE_{\theta_{MVU}} + \theta_0^2}$$

(h) Because we do not know the value of θ_0. If we knew the real value of θ_0 we would not need to try finding estimators.

Exercise 3

(a) $\left(\sum_{n=1}^{N} x_n x_n^T\right)\theta = \sum_{n=1}^{N} y_n x_n \Rightarrow N\hat{\theta}_{LS} = \sum_{n=1}^{N} y_n \Rightarrow \hat{\theta}_{LS} = \frac{1}{N}\sum_{n=1}^{N} y_n$

(b) $E[y_n] = E[\theta_0 + n] \Rightarrow E[y_n] = E[\theta_0] + E[n]$

But the nois has a mean equal to zero, thus:

$E[y_n] = E[\theta_0] = \theta_0$

Therefore, y_n is an unbiased estimator of θ_0

(c) $\bar{y} = \frac{1}{N}\sum_{n=1}^{N} y_n \Rightarrow E[\bar{y}] = \frac{1}{N} E[\sum_{n=1}^{N} y_n] \Rightarrow E[\bar{y}] = \frac{1}{N}\cdot N\theta_0 \Rightarrow$

$\Rightarrow E[\bar{y}] = \theta_0$ Thus, $\bar{y}$ is an unbiased estimator of θ_0

(e) $\left(\sum_{n=1}^{\tilde{N}}(x_n x_n^T) + \lambda I\right)\hat{\theta} = \sum_{n=1}^{\tilde{N}} y_n x_n$

But $x_n = 1$, thus:

$$(N+\lambda)\hat{\theta}_{ridge} = \sum_{n=1}^{\tilde{N}} y_n \Rightarrow \hat{\theta}_{ridge} = \frac{\sum_{n=1}^{\tilde{N}} y_n}{N+\lambda}$$

(f) We know that $\hat{\theta}_{LS} = \frac{1}{N}\sum_{n=1}^{\tilde{N}} y_n \Rightarrow \sum_{n=1}^{\tilde{N}} y_n = N\hat{\theta}_{LS}$ or $\sum_{n=1}^{\tilde{N}} y_n = N\tilde{\theta}_{MVU}$

from (e): $\hat{\theta}_{ridge} = \theta = \frac{N\hat{\theta}_{MVU}}{N+\lambda}$

(g) $E[\hat{\theta}] = E\left[\frac{N}{N+\lambda}\hat{\theta}_{MVU}\right] \xrightarrow{E(\hat{\theta}_{MVU}) = \theta_0} E[\hat{\theta}] = \frac{N}{N+\lambda}\theta_0 \Rightarrow$

$\Rightarrow E[\hat{\theta}] = \frac{N}{N+\lambda}\theta_0$ For $\hat{\theta}$ to be unbiased estimator it needs to hold that $E[\hat{\theta}] = \theta_0 \Rightarrow \frac{N}{N+\lambda} = 1 \Rightarrow \lambda = 0$

But $\lambda > 0$, thus $\hat{\theta}$ is biased estimator of θ_0

(h) $|\hat{\theta}| = \left|\frac{N}{N+\lambda}\right| |\hat{\theta}_{MVU}|$ For $|\hat{\theta}|$ to be smaller than $|\hat{\theta}_{MVU}|$ it needs to hold that

$\left|\frac{N}{N+\lambda}\right| < 1$. Indeed for $\lambda > 0$ $\left|\frac{N}{N+\lambda}\right| < 1$

Thus, $|\hat{\theta}| < |\hat{\theta}_{MVU}|$

(i) $\hat{\theta}_b = (1+a)\hat{\theta}_{MVU} \Rightarrow 1+a = \frac{N}{N+\lambda} \Rightarrow a = \frac{-\lambda}{N+\lambda}$

$a > \frac{-2MSE(\theta_{MVU})}{MSE(\theta_{MVU}) + \theta_0^2} \Rightarrow -\frac{\lambda}{N+\lambda} > \frac{-2MSE(\theta_{MVU})}{MSE(\theta_{MVU}) + \theta_0^2} \Rightarrow$

Sofia Baltzi
29/01/2020

$\Rightarrow \lambda < \frac{-2N MSE(\theta_{MVU})}{MSE(\theta_{MVU}) + \theta_0^2} - \frac{2\lambda MSE(\theta_{MVU})}{MSE(\theta_{MVU}) + \theta_0^2} \Rightarrow$

$\Rightarrow \lambda < \frac{2MSE(\theta_{MVU})}{\theta_0^2 - MSE(\theta_{MVU})}$, also $\lambda > 0$