Notes - Theoretical exercises 1

November 3, 2021

Notice that deadlines for homework are on Sunday 22:00!

Exercise 1.1

a)

- $(AB)^T = B^T A^T$ and $(A^T)^T = A$. Thus
$$x^T A = ((x^T A)^T)^T = (A^T x)^T.$$
- Remember that $(AB)C = A(BC)$. Thus $(x^T A)x = x^T(Ax)$.

Exercise 1.2

a)

$$\begin{aligned} f(\beta) = \varepsilon^T\varepsilon = (y - X\beta)^T(y - X\beta) &= (y^T - \beta^T X^T)(y - X\beta) \\ &= y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta. \end{aligned}$$

Remember that $x^T y = y^T x$. Then $y^T X\beta = (X\beta)^T y = \beta^T X^T y$ and thus

$$f(\beta) = y^T y - 2\beta^T X^T y + \beta^T X^T X\beta.$$

Now

$$\begin{aligned} &\frac{\partial y^T y}{\partial \beta} = 0, \\ &\frac{\partial(-2\beta^T X^T y)}{\partial \beta} = -2\frac{\partial \beta^T (X^T y)}{\partial \beta} \overset{1.1b)}{=} -2(X^T y)^T = -2y^T X \quad \text{and} \\ &\frac{\partial \beta^T (X^T X)\beta}{\partial \beta} \overset{1.1f)}{=} \beta^T(X^T X + \underbrace{(X^T X)^T}_{=X^T X}) = 2\beta^T X^T X. \end{aligned}$$

Then we get

$$f'(\beta) = -2y^T X + 2\beta^T X^T X.$$

By setting $f'(\beta) = 0$ we get

$$\begin{aligned} &\beta^T X^T X = y^T X \\ &\Rightarrow X^T X \beta = X^T y \\ &\Rightarrow \beta = (X^T X)^{-1} X^T y. \end{aligned}$$

In order to prove that $b = (X^T X)^{-1} X^T y$ is minimum we have to prove that $f''(\beta) = X^T X$ is symmetric positive definite. Let $a \in \mathbb{R}^{k+1} \setminus \{0\}$. Denote

$$\begin{aligned} &a = (a_1 \dots a_{k+1})^T \quad \text{and} \\ &X = (x_1 \dots x_{k+1}), \end{aligned}$$

where $x_i \in \mathbb{R}^{k+1}$ are the column vectors of X. We have that vector $\sum_{i=1}^{k+1} a_i x_i$ is always **not** equal to zero since vectors x_i are linearly independent. Then

$$a^T (X^T X) a = (Xa)^T X a = \|Xa\|^2 = \left\| \sum_{i=1}^{k+1} a_i x_i \right\|^2 > 0.$$

Thus $X^T X$ is positive definite.

b)

$$b = (X^T X)^{-1} X^T (X\beta + \varepsilon) = \underbrace{(X^T X)^{-1} X^T X}_{=I} \beta + (X^T X)^{-1} X^T \varepsilon = \beta + (X^T X)^{-1} X^T \varepsilon$$

Thus

$$\mathbb{E}(b) = \mathbb{E}(\beta) + (X^T X)^{-1} X^T \, \mathbb{E}(\varepsilon) = \beta.$$

c) By part b),

$$b - \mathbb{E}(b) = b - \beta = \beta + (X^T X)^{-1} X^T \varepsilon - \beta = (X^T X)^{-1} X^T \varepsilon$$

Also, notice that

$$\mathbb{E}(\varepsilon \varepsilon^T) = \mathbb{E}\left((\varepsilon - \underbrace{\mathbb{E}\,\varepsilon}_{=0})(\varepsilon - \underbrace{\mathbb{E}\,\varepsilon}_{=0})^T \right) = \operatorname{cov}(\varepsilon) = \sigma^2 I,$$

and remember that $(A^{-1})^T = (A^T)^{-1}$. Then

$$
\begin{aligned}
\operatorname{cov}(b) &= \mathbb{E}\left((b-\mathbb{E}\,b)(b-\mathbb{E}\,b)^T\right) = \mathbb{E}(((X^TX)^{-1}X^T\varepsilon)((X^TX)^{-1}X^T\varepsilon)^T) \\
&= \mathbb{E}((X^TX)^{-1}X^T\varepsilon\varepsilon^TX(X^TX)^{-1}) = (X^TX)^{-1}X^T\,\mathbb{E}(\varepsilon\varepsilon^T)X(X^TX)^{-1} \\
&= (X^TX)^{-1}X^T(\sigma^2 I)X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}\underbrace{(X^TX)(X^TX)^{-1}}_{=I} \\
&= \sigma^2(X^TX)^{-1}
\end{aligned}
$$

Exercise 1.3 (Homework)

a)

- Use properties of projection matrices. That is,

$$M^T = M \text{ and } M^2 = M.$$

- Use linearity of expectation.
- Write e in terms of matrix M.
- First calculate $\operatorname{cov}(y)$. It will be useful in calculating $\operatorname{cov}(e)$.

b)

- First compute $\mathbb{E}(e)$. Then find relation between $\operatorname{trace}(\operatorname{cov}(e))$ and $\mathbb{E}(\sum_{i=1}^n e_i^2)$.
- Let $A \in \mathbb{R}^{n\times n}$. Then trace is defined as $\operatorname{trace}(A) = \sum_{i=1}^n a_{ii}$
- $\operatorname{trace}(cA) = c\operatorname{trace}(A), c \in \mathbb{R}$.
- For square matrices, Rank is equal to the number of nonzero eigenvalues.