1. Theoretical exercises

Demo exercises

1.1 Let $\boldsymbol{A}$ be a real valued 2×2-matrix and let $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^2$,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

(a) Write explicitly $\boldsymbol{A}^\top$, $\boldsymbol{A}^\top \boldsymbol{A}$ $\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x}^\top \boldsymbol{A}$ and $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$.
Solution.

$$\boldsymbol{A}^\top = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}, \qquad \boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} a_{11}^2 + a_{21}^2 & a_{11}a_{12} + a_{21}a_{22} \\ a_{11}a_{12} + a_{21}a_{22} & a_{12}^2 + a_{22}^2 \end{bmatrix},$$

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}, \qquad \boldsymbol{x}^\top \boldsymbol{A} = \begin{bmatrix} a_{11}x_1 + a_{21}x_2 & a_{12}x_1 + a_{22}x_2, \end{bmatrix}$$

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2$$

(b) Verify that $\frac{\partial(\boldsymbol{b}^\top \boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial(\boldsymbol{x}^\top \boldsymbol{b})}{\partial \boldsymbol{x}} = \boldsymbol{b}^\top$.
Solution. Since,

$$\boldsymbol{b}^\top \boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{b} = b_1x_1 + b_2x_2,$$

and

$$\begin{bmatrix} \frac{\partial}{\partial x_1}(b_1x_1 + b_2x_2) & \frac{\partial}{\partial x_2}(b_1x_1 + b_2x_2) \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \end{bmatrix} = \boldsymbol{b}^\top,$$

which proves the claim. Note that the result can be generalized for $\mathbb{R}^n$-vectors.

(c) Verify that $\frac{\partial(\boldsymbol{x}^\top \boldsymbol{x})}{\partial \boldsymbol{x}} = 2\boldsymbol{x}^\top$
Solution. The claim follows by taking the derivatives from,

$$\boldsymbol{x}^\top \boldsymbol{x} = x_1^2 + x_2^2.$$

Note that the result can be generalized for $\mathbb{R}^n$-vectors.

(d) Verify that $\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}$.
Solution.

$$\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1}(a_{11}x_1 + a_{12}x_2) & \frac{\partial}{\partial x_2}(a_{11}x_1 + a_{12}x_2) \\ \frac{\partial}{\partial x_1}(a_{21}x_1 + a_{22}x_2) & \frac{\partial}{\partial x_2}(a_{21}x_1 + a_{22}x_2) \end{bmatrix} = \boldsymbol{A}.$$

Note that the result can be generalized for $\mathbb{R}^n$-vectors and $\mathbb{R}^{m \times n}$-matrices.

(e) Verify that $\frac{\partial(\boldsymbol{x}^\top A)}{\partial \boldsymbol{x}} = \boldsymbol{A}^\top$.
Solution.

$$\begin{aligned} \frac{\partial(\boldsymbol{x}^\top \boldsymbol{A})}{\partial \boldsymbol{x}} &= \frac{\partial}{\partial \boldsymbol{x}} \begin{bmatrix} a_{11}x_1 + a_{21}x_2 & a_{12}x_1 + a_{22}x_2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial x_1}(a_{11}x_1 + a_{21}x_2) & \frac{\partial}{\partial x_2}(a_{11}x_1 + a_{21}x_2) \\ \frac{\partial}{\partial x_1}(a_{12}x_1 + a_{22}x_2) & \frac{\partial}{\partial x_2}(a_{12}x_1 + a_{22}x_2) \end{bmatrix} = \boldsymbol{A}^\top. \end{aligned}$$

Note that the result can be generalized for $\mathbb{R}^n$-vectors and $\mathbb{R}^{n \times n}$-matrices.

(f) Verify that $\frac{\partial(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{x}^\top(\boldsymbol{A}+\boldsymbol{A}^\top)$.

Solution.

$$\begin{aligned}\frac{\partial(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} &= \left[\tfrac{\partial}{\partial x_1}(a_{11}x_1^2 + (a_{12}+a_{21})x_1x_2 + a_{22}x_2^2) \quad \tfrac{\partial}{\partial x_2}(a_{11}x_1^2 + (a_{12}+a_{21})x_1x_2 + a_{22}x_2^2)\right] \\ &= \left[2a_{11}x_1 + (a_{12}+a_{21})x_2 \quad (a_{12}+a_{21})x_1 + 2a_{22}x_2\right] \\ &= \left[a_{11}x_1 + a_{12}x_2 \quad a_{21}x_1 + a_{22}x_2\right] + \left[a_{11}x_1 + a_{21}x_2 \quad a_{12}x_1 + a_{22}x_2\right] \\ &= \boldsymbol{x}^\top\boldsymbol{A}^\top + \boldsymbol{x}^\top\boldsymbol{A} = \boldsymbol{x}^\top(\boldsymbol{A}+\boldsymbol{A}^\top).\end{aligned}$$

Note that the result can be generalized for $\mathbb{R}^n$-vectors and $\mathbb{R}^{n\times n}$-matrices.

Remark: consider a vector valued function $f : \mathbb{R}^n \to \mathbb{R}^m$, that is smooth enough,

$$f(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}, \quad \boldsymbol{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n.$$

The derivative of the function is

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_1(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} \\ \frac{\partial f_2(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_2(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\boldsymbol{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_m(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix}.$$

1.2 Consider the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{y}, \boldsymbol{\varepsilon} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n\times(k+1)}$ and $\boldsymbol{\beta} \in \mathbb{R}^{k+1}$. Let the standard assumptions (i)–(v), given in the lecture slides, hold.

a) Show that the least squares (LS) estimator for the vector $\boldsymbol{\beta}$ is $\mathbf{b} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$.

b) Show that $\mathbf{b}$ is unbiased, that is, $\mathbb{E}[\mathbf{b}] = \boldsymbol{\beta}$.

c) Show that $\mathrm{Cov}(\mathbf{b}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$.

Solution.

a) The sum of squares is,

$$f(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^\top\mathbf{y} - 2\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{y} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}.$$

The LS-estimator for the vector $\boldsymbol{\beta}$ is obtained by minimizing the sum of squares $f(\boldsymbol{\beta})$ with respect to the vector $\boldsymbol{\beta}$. By differentiating $f(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and setting the derivative equal to zero, we obtain,

$$f'(\boldsymbol{\beta}) = -2\mathbf{y}^\top\mathbf{X} + 2\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X} = \mathbf{0}.$$

By the standard assumptions, we have that $\text{rank}(\mathbf{X}) = k+1$, which implies that the matrix $\mathbf{X}^\top\mathbf{X}$ is nonsingular and $\boldsymbol{\beta}$ is solvable. The solution,

$$\mathbf{b} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y},$$

minimizes the function $f(\boldsymbol{\beta})$, since $\mathbf{X}^\top\mathbf{X}$ is always a positive definite matrix and,

$$f''(\boldsymbol{\beta}) = 2\mathbf{X}^\top\mathbf{X}.$$

b) Note that,

$$\mathbf{b} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}.$$

Since the vector $\boldsymbol{\beta}$ and the matrix $\mathbf{X}$ are non-random, it follows that,

$$\mathbb{E}[\mathbf{b}] = \mathbb{E}[\boldsymbol{\beta}] + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbb{E}[\boldsymbol{\varepsilon}] = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{0} = \boldsymbol{\beta}.$$

c) Using part (b), we obtain,

$$\mathbf{b} - \mathbb{E}[\mathbf{b}] = \mathbf{b} - \boldsymbol{\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}.$$

Then, since $\mathbf{X}$ is non-random,

$$\begin{aligned}
\text{Cov}(\mathbf{b}) &= \mathbb{E}[(\mathbf{b}-\mathbb{E}[\mathbf{b}])(\mathbf{b}-\mathbb{E}[\mathbf{b}])^\top] \\
&= E[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}] \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top]\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}.
\end{aligned}$$

Homework

1.3 Consider the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{y}, \boldsymbol{\varepsilon} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n\times(k+1)}$ and $\boldsymbol{\beta} \in \mathbb{R}^{k+1}$. Let the standard assumptions (i)–(v), given in the lecture slides, hold. Let $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ and recall that $\text{rank}(\mathbf{M}) = n - (k+1)$.

a) Let $\mathbf{e}$ be the estimated residual vector, that is, $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$. Use the results obtained in previous exercises to show that

$$\text{Cov}(\mathbf{e}) = \sigma^2\mathbf{M}.$$

b) Use previous exercises and part (a), and show that,

$$s^2 = \frac{1}{n-k-1}\sum_{i=1}^{n} e_i^2,$$

is an unbiased estimator for $\text{Var}[\varepsilon_i] = \sigma^2$, that is, show that $\mathbb{E}[s^2] = \sigma^2$.
Hint. (1) The trace of a square matrix equals the sum of the corresponding eigenvalues, and, (2) the eigenvalues of an idempotent matrix are either 0 or 1. (You are not expected to prove results (1)-(2) in this exercise.)