Marcelo Almora Rios
Math189R
Homework 1
June 13, 2020

Feel free to work with other students, but make sure you write up the homework and code on your own (no copying homework *or* code; no pair programming). Feel free to ask students or instructors for help debugging code or whatever else, though. The starter code for problem 2 part c and d can be found under the Resource tab on course website.

*Note:* You need to create a Github account for submission of the coding part of the homework. Please create a repository on Github to hold all your code and include your Github account username as part of the answer to problem 2.

**1** (**Linear Transformation**) Let $\mathbf{y} = A\mathbf{x} + \mathbf{b}$ be a random vector. show that expectation is linear:

$$\mathbb{E}[\mathbf{y}] = \mathbb{E}[A\mathbf{x} + \mathbf{b}] = A\mathbb{E}[\mathbf{x}] + \mathbf{b}.$$

Also show that

$$\text{cov}[\mathbf{y}] = \text{cov}[A\mathbf{x} + \mathbf{b}] = A\text{cov}[\mathbf{x}]A^\top = A\mathbf{\Sigma}A^\top.$$

Suppose the cumulative distribution function of the $N$-dimensional random vector $\mathbf{y}$ admits a density $\mathbb{F}$. Then

$$\mathbb{E}[\mathbf{y}] = \int_{\mathbb{R}} \left(A\mathbf{x} + \mathbf{b}\right) \mathbb{F}(x)\,\mathrm{d}x = A\int_{\mathbb{R}} \mathbf{x}\,\mathbb{F}(x)\,\mathrm{d}x + \mathbf{b}\int_{\mathbb{R}} \mathbb{F}(x)\,\mathrm{d}x$$

by linearity of the Lebesgue integral. Because

$$\int_{\mathbb{R}} \mathbb{F}(x)\,\mathrm{d}x = 1,$$

the claim is proven true.

Observe that by matrix transpose properties and the linearity of the expectation,

$$\begin{aligned}
\text{cov}[\mathbf{y}] &= \mathbb{E}\Big[\big(A\mathbf{x} + \mathbf{b} - \mathbb{E}[A\mathbf{x} + \mathbf{b}]\big)\big(A\mathbf{x} + \mathbf{b} - \mathbb{E}[A\mathbf{x} + \mathbf{b}]\big)^T\Big] \\
&= \mathbb{E}\Big[\big(A\mathbf{x} - A\mathbb{E}[\mathbf{x}]\big)\big(A\mathbf{x} - A\mathbb{E}[\mathbf{x}]\big)^T\Big] \\
&= A\mathbb{E}\Big[\big(\mathbf{x} - \mathbb{E}[\mathbf{x}]\big)\big(\mathbf{x} - \mathbb{E}[\mathbf{x}]\big)^T\Big]A^T \\
&= A\text{cov}[\mathbf{x}]A^T.
\end{aligned}$$

Then,

$$
\begin{aligned}
\operatorname{cov}[\mathbf{x}] &= \mathbb{E}\Big[\big(\mathbf{x}-\mathbb{E}[\mathbf{x}]\big)\big(\mathbf{x}-\mathbb{E}[\mathbf{x}]\big)^T\Big] \\
&= \mathbb{E}\left[\begin{bmatrix} X_1-\mu_1 \\ \cdots \\ X_N-\mu_N \end{bmatrix} \cdot \begin{bmatrix} X_1-\mu_1 \\ \cdots \\ X_N-\mu_N \end{bmatrix}^T\right] \\
&= \mathbb{E}\left[\begin{bmatrix} (X_1-\mu_1)(X_1-\mu_1) & (X_1-\mu_1)(X_2-\mu_2) & \cdots & (X_1-\mu_1)(X_N-\mu_N) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)(X_2-\mu_2) & \cdots & (X_2-\mu_2)(X_N-\mu_N) \\ \vdots & \vdots & \ddots & \vdots \\ (X_N-\mu_N)(X_1-\mu_1) & (X_N-\mu_N)(X_2-\mu_2) & \cdots & (X_N-\mu_N)(X_N-\mu_N) \end{bmatrix}\right] \\
&= \begin{bmatrix} E\big[(X_1-\mu_1)(X_1-\mu_1)\big] & E\big[(X_1-\mu_1)(X_2-\mu_2)\big] & \cdots & E\big[(X_1-\mu_1)(X_N-\mu_N)\big] \\ E\big[(X_2-\mu_2)(X_1-\mu_1)\big] & E\big[(X_2-\mu_2)(X_2-\mu_2)\big] & \cdots & E\big[(X_2-\mu_2)(X_N-\mu_N)\big] \\ \vdots & \vdots & \ddots & \vdots \\ E\big[(X_N-\mu_N)(X_1-\mu_1)\big] & E\big[(X_N-\mu_N)(X_2-\mu_2)\big] & \cdots & E\big[(X_N-\mu_N)(X_N-\mu_N)\big] \end{bmatrix} \\
&= \mathbf{\Sigma}
\end{aligned}
$$

as desired.

$\blacksquare$

**2** Given the dataset $\mathcal{D} = \{(x, y)\} = \{(0,1), (2,3), (3,6), (4,8)\}$

(a) Find the least squares estimate $y = \boldsymbol{\theta}^\top \mathbf{x}$ by hand using Cramer's Rule.

(b) Use the normal equations to find the same solution and verify it is the same as part (a).

(c) Plot the data and the optimal linear fit you found.

(d) Find randomly generate 100 points near the line with white Gaussian noise and then compute the least squares estimate (using a computer). Verify that this new line is close to the original and plot the new dataset, the old line, and the new line.

(a) We wish to find $\theta_0$ and $\theta_1$ such that

$$y^{(i)} = \theta_0 + \theta_1 x^{(i)} \tag{1}$$

for all $i = 1, \ldots, 4$, where $(x^{(i)}, y^{(i)}) \in \mathcal{D}$. We can express equation (1) given dataset $\mathcal{D}$ in matrix notation as

$$\begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}. \tag{2}$$

Let

$$\mathbf{X} \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \tag{3}$$

To use Cramer's Rule, we need square matrices. Let's left-multiply both sides by $\mathbf{X}^T$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}. \tag{4}$$

Evaluating,

$$\begin{bmatrix} 18 \\ 56 \end{bmatrix} = \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}. \tag{5}$$

Thus,

$$\theta_0 = \frac{\det \begin{bmatrix} 18 & 9 \\ 56 & 29 \end{bmatrix}}{\det \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}} = \frac{18}{35} \qquad \text{and} \qquad \theta_1 = \frac{\det \begin{bmatrix} 4 & 18 \\ 9 & 56 \end{bmatrix}}{\det \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}} = \frac{62}{35}, \tag{6}$$

via Cramer's Rule.

(b) According to the normal equations:

$$\boldsymbol{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{7}$$

See that

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}^{-1} = \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}^{-1} = \frac{1}{35}\begin{bmatrix} 29 & -9 \\ -9 & 4 \end{bmatrix}. \tag{8}$$

So

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \frac{1}{35}\begin{bmatrix} 29 & -9 \\ -9 & 4 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 29 & 11 & 2 & -7 \\ -9 & -1 & 3 & 7 \end{bmatrix}. \tag{9}$$

Which means that

$$\boldsymbol{\theta} = \frac{1}{35}\begin{bmatrix} 29 & 11 & 2 & -7 \\ -9 & -1 & 3 & 7 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 18 \\ 62 \end{bmatrix}, \tag{10}$$

as shown in part (a).

(c)

■