Name: Kieran Saucedo
MATH 179 SP24
Homework 1
Due Monday 1/29

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.

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

$$E[A\mathbf{x} + \mathbf{b}] = AE[\mathbf{x}] + \mathbf{b}$$

Also show that

$$cov[\mathbf{y}] = cov[A\mathbf{x} + \mathbf{b}] = Acov[\mathbf{x}]A^T = A\sum A^T$$

1. **(a)** Showing that expectation is a linear operator.
   First, note that:

$$E[X] = \underbrace{\sum_X xP(x)}_{\text{discrete}} = \underbrace{\int_X xf(x)\,dx}_{\text{continuous}}$$

   depending on whether we're referring to a discrete or continuous random variable $X$.

   In the first case, assuming $E[X] = \sum_X xP(x)$ and $\mathbf{y} = A\mathbf{x} + \mathbf{b}$:

$$\begin{aligned}
E[\mathbf{y}] &= E[A\mathbf{x} + \mathbf{b}] \\
&= \sum_X (A\mathbf{x} + \mathbf{b})P(x) \\
&= \sum_X (A\mathbf{x})P(x) + \sum_X (\mathbf{b})P(x) \\
&= A\sum_X (\mathbf{x})P(x) + \mathbf{b}\sum_X (1)P(x) \\
&= AE[\mathbf{x}] + \mathbf{b}
\end{aligned}$$

   Meaning $E[\mathbf{y} = A\mathbf{x} + \mathbf{b}] = AE[\mathbf{x}] + \mathbf{b}$, as desired.

   In the second case, assuming $E[X] = \int_X xf(x)\,dx$ and $\mathbf{y} = A\mathbf{x} + \mathbf{b}$:

$$
\begin{aligned}
E[\mathbf{y}] &= E[A\mathbf{x} + \mathbf{b}] \\
&= \int_X (A\mathbf{x} + \mathbf{b}) f(x)\,dx \\
&= \int_X (A\mathbf{x}) f(x)\,dx + \int_X (\mathbf{b}) f(x)\,dx \\
&= (A) \int_X (\mathbf{x}) f(x)\,dx + (\mathbf{b}) \int_X (1) f(x)\,dx \\
&= AE[\mathbf{x}] + \mathbf{b}
\end{aligned}
$$

Meaning $E[\mathbf{y} = A\mathbf{x} + \mathbf{b}] = AE[\mathbf{x}] + \mathbf{b}$, as desired.

In either case, expectation is a linear operator and $E[\mathbf{y} = A\mathbf{x} + \mathbf{b}] = AE[\mathbf{x}] + \mathbf{b}$.

2. **(b)** Showing that $cov[\mathbf{y}] = A \sum A^T$:

   First, note that:

$$
cov[\mathbf{y}] = \sum = E[(x - E[x])(x - E[x])^T]
$$

   Therefore:

$$
\begin{aligned}
cov[\mathbf{y}] &= cov[A\mathbf{x} + \mathbf{b}] \\
&= E[(A\mathbf{x} + \mathbf{b} - E[A\mathbf{x} + \mathbf{b}])(A\mathbf{x} + \mathbf{b} - E[A\mathbf{x} + \mathbf{b}])^T] \\
&= E[(A\mathbf{x} + \mathbf{b} - AE[\mathbf{x}] - \mathbf{b})(A\mathbf{x} + \mathbf{b} - AE[\mathbf{x}] - \mathbf{b})^T] \\
&= E[(A\mathbf{x} - AE[\mathbf{x}])(A\mathbf{x} - AE[\mathbf{x}])^T] \\
&= E[A(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T A^T] \\
&= AE[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T]A^T \\
&= A \sum A^T
\end{aligned}
$$

   Meaning $cov[A\mathbf{x} + \mathbf{b}] = A \sum A^T$, as desired.

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

1. Find the least squares estimate $y = \mathbf{b}^T\mathbf{x}$ by hand using Cramer's Rule.
2. Use the normal equations to find the same solution and verify it is the same as part (a).
3. Plot the data and the optimal linear fit you found.
4. 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.

**(1)** Using Cramer's Rule to find the least squares estimate $y = \mathbf{b}^T\mathbf{x}$ by hand:

First, note that (for our 4 data points):

- $\sum_{i=1}^{4}(x_i \cdot y_i) = (0 \cdot 1) + (2 \cdot 3) + (3 \cdot 6) + (4 \cdot 8) = 56$
- $\sum_{i=1}^{4}(x_i) = 0 + 2 + 3 + 4 = 9$
- $\sum_{i=1}^{4}(y_i) = 1 + 3 + 6 + 8 = 18$
- $\sum_{i=1}^{4}(x_i)^2 = 0^2 + 2^2 + 3^2 + 4^2 = 0 + 4 + 9 + 16 = 29$

Accordingly, we calculate:

$$m = \frac{(n \cdot \sum_{i=1}^{4}(x_i \cdot y_i)) - (\sum_{i=1}^{4} x_i \cdot \sum_{i=1}^{4} y_i)}{(n \cdot \sum_{i=1}^{4}(x_i)^2) - (\sum_{i=1}^{4} x_i)^2} = \frac{(4 \cdot 56) - (9 \cdot 18)}{(4 \cdot 29) - (9)^2} = \frac{62}{35}$$

$$b = \frac{(\sum_{i=1}^{4}(x_i)^2) \cdot (\sum_{i=1}^{4} y_i) - (\sum_{i=1}^{4} x_i) \cdot (\sum_{i=1}^{4}(x_i \cdot y_i))}{(n \cdot (\sum_{i=1}^{4}(x_i)^2) - (\sum_{i=1}^{4} x_i)^2} = \frac{(29 \cdot 18) - (9 \cdot 56)}{(4 \cdot 29) - (9)^2} = \frac{18}{35}$$

Which yields a least square estimate:

$$y = \mathbf{b}^T\mathbf{x} = \begin{bmatrix} \frac{18}{35} \\ \frac{62}{35} \end{bmatrix} \cdot \begin{bmatrix} 1 & x \end{bmatrix} = \frac{18}{35} + \frac{62}{35}x$$

**(2)** Next we use the normal equation $\mathbf{b^T} = (X^T X)^{-1} \cdot X^T \mathbf{y}$ to derive $\mathbf{b^T}$:

First, note that:

- $X = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$
- $X^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix}$
- $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$

Such that:

$$\begin{aligned} \mathbf{b^T} &= (X^T X)^{-1} \cdot X^T \mathbf{y} \\ &= \left( \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \right)^{-1} \cdot \left( \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} \right) \\ &= \left( \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 18 \\ 56 \end{bmatrix} \\ &= \begin{bmatrix} \frac{29}{35} & \frac{-9}{35} \\ \frac{-9}{35} & \frac{4}{35} \end{bmatrix} \cdot \begin{bmatrix} 18 \\ 56 \end{bmatrix} \\ &= \begin{bmatrix} \frac{18}{35} \\ \frac{62}{35} \end{bmatrix} \end{aligned} \tag{1}$$

$$\begin{aligned} \mathbf{b^T} &= (X^T X)^{-1} \cdot X^T \mathbf{y} \\ &= \left( \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \right)^{-1} \cdot \left( \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} \right) \\ &= \left( \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 18 \\ 56 \end{bmatrix} \\ &= \begin{bmatrix} \frac{29}{35} & \frac{-9}{35} \\ \frac{-9}{35} & \frac{4}{35} \end{bmatrix} \cdot \begin{bmatrix} 18 \\ 56 \end{bmatrix} \\ &= \begin{bmatrix} \frac{18}{35} \\ \frac{62}{35} \end{bmatrix} \end{aligned} \tag{2}$$