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

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

Intuitively, since *A* and *b* are constants, the expected value of them stays constant.

$$E[y] = E[Ax + b] = E[Ax] + b = AE[x] + b$$

We can write $y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = Ax + b = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & \ddots & & \\ \cdots & & & \\ A_{m1} & & & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix} = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 + \cdots A_{1n}x_n + b_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots A_{2n}x_n + b_2 \\ \cdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots A_{mn}x_n + b_n \end{bmatrix}$

Thus,

$$E[y] = \begin{bmatrix} E(y_1) \\ E(y_2) \\ \cdots \\ E(y_2) \end{bmatrix} = E[Ax + b] = E\left[\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & \ddots & & \\ \cdots & & & \\ A_{m1} & & & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix}\right] = \begin{bmatrix} E[A_{11}x_1 + A_{12}x_2 + \cdots A_{1n}x_n + b_1] \\ E[A_{21}x_1 + A_{22}x_2 + \cdots A_{2n}x_n + b_2] \\ \cdots \\ E[A_{m1}x_1 + A_{m2}x_2 + \cdots A_{mn}x_n + b_n] \end{bmatrix}$$

Since the entries in **b** are constant, their expected value remains the same, so:

$$E[y] = \begin{bmatrix} E[A_{11}x_1 + A_{12}x_2 + \cdots A_{1n}x_n] + b_1 \\ E[A_{21}x_1 + A_{22}x_2 + \cdots A_{2n}x_n] + b_2 \\ \cdots \\ E[A_{m1}x_1 + A_{m2}x_2 + \cdots A_{mn}x_n] + b_n \end{bmatrix} = \begin{bmatrix} E[A_{11}x_1 + A_{12}x_2 + \cdots A_{1n}x_n] \\ E[A_{21}x_1 + A_{22}x_2 + \cdots A_{2n}x_n] \\ \cdots \\ E[A_{m1}x_1 + A_{m2}x_2 + \cdots A_{mn}x_n] \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix} \quad \blacksquare$$

Similarly, since the entries in *A* are also constant,

$$E[\mathbf{y}] = \begin{bmatrix} A_{11}E[X_1] + A_{12}E[X_2] + \ldots + A_{1n}E[X_n] \\ A_{21}E[X_1] + A_{22}E[X_2] + \ldots + A_{2n}E[X_n] \\ \cdots \\ A_{m1}E[X_1] + A_{m2}E[X_2] + \ldots + A_{mn}E[X_n] \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix}$$

Notice this can be decomposed to:

$$E[\mathbf{y}] = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & \ddots & & \\ \cdots & & & \\ A_{m1} & & & A_{mn} \end{bmatrix} \begin{bmatrix} E[X_1] \\ E[X_2] \\ \cdots \\ E[X_n] \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix} = A\,E[\mathbf{x}] + b$$

To prove the linearity of covariance: $\mathrm{cov}[\mathbf{y}] = \mathrm{cov}[A\mathbf{x} + b]$

Thus we can write $E\left[(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}])^T\right] = E\left[\left((A\mathbf{x}+b) - E[A\mathbf{x}+b]\right)\left((A\mathbf{x}+b) - E[A\mathbf{x}+b]\right)^T\right]$

Substituting our response from the previous part: $= \left(A\mathbf{x} + b - (A\,E[\mathbf{x}] + b)\right)\left(A\mathbf{x} + b - (A\,E[\mathbf{x}] + b)\right)^T$

$$= \left(A\mathbf{x} - A\,E[\mathbf{x}]\right)\left(A\mathbf{x} - A\,E[\mathbf{x}]\right)^T$$

And factoring out *A*, $= A\left(\mathbf{x} - E[\mathbf{x}]\right)\left(\mathbf{x} - E[\mathbf{x}]\right)^T A^T$

$$= A\,\mathrm{cov}[\mathbf{x}]\,A^T = A\Sigma A^T$$

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) As Cramer's Rule derives, to minimize the error,

$$m = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \quad b = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Calculating the values from our dataset: $\sum_{i=1}^{4} x_i = 0 + 2 + 3 + 4 = 9$, $\sum_{i=1}^{4} y_i = 1 + 3 + 6 + 8 = 18$,

$\sum_{i=1}^{4} x_i^2 = 0 + 4 + 9 + 16 = 29$, $\sum_{i=1}^{4} x_i y_i = 0 + 6 + 18 + 32 = 56$

So, $$m = \frac{4(56) - (9)(18)}{4(29) - 9^2} = \frac{62}{35}, \quad b = \frac{(29)(18) - (9)(56)}{4(29) - 9^2} = \frac{18}{35}$$

Thus, $\theta = \begin{bmatrix} \frac{18}{35} \\ \frac{62}{35} \end{bmatrix}$, so we can express $y = \begin{bmatrix} \frac{18}{35} \\ \frac{62}{35} \end{bmatrix}^T X$

∎

b) We are given $X = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$, $\vec{y} = \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$.

Also recall the normal equation: $\theta = (X^T X)^{-1} X^T \vec{y}$

We confirm $$\left(\begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}^T \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$$

$$= \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} \frac{18}{35} \\ \frac{62}{35} \end{bmatrix}$$

(Calculations done using Symbolab)

c) See code and image in GitHub submission, hw1pr2c.png

d) See code and image in GitHub submission, hw1pr2d.png