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

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

Show expectation is linear:
Suppose the density function is $\mathbf{P}(\mathbf{x})$, and that it is continuous. Then,

$$\begin{aligned}
\mathbb{E}[A\mathbf{x} + \mathbf{b}] &= \int (A\mathbf{x} + \mathbf{b}) \cdot \mathbf{P} d\mathbf{x} \\
&= \int (A\mathbf{x}\mathbf{P} + \mathbf{b}\mathbf{P}) d\mathbf{x} \\
&= A \int \mathbf{x}\mathbf{P} d\mathbf{x} + \mathbf{b} \int \mathbf{P} d\mathbf{x} \\
&= A\mathbb{E}[\mathbf{x}] + \mathbf{b}
\end{aligned}$$

■

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

By definition of the covariance matrix on random vector,

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

■

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

(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) Using Cramer's Rule,

$$\begin{aligned} m &= \frac{n\sum_{i=1}^{n}(x_i y_i) - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \\ &= \frac{4\sum_{i=1}^{4}(x_i y_i) - \sum_{i=1}^{4} x_i \sum_{i=1}^{4} y_i}{4\sum_{i=1}^{4} x_i^2 - (\sum_{i=1}^{4} x_i)^2} \\ &= \frac{4 * 56 - 9 * 18}{4 * 29 - 81} \\ &= \frac{62}{35} \end{aligned}$$

$$\begin{aligned} b &= \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n}(x_i y_i)}{n\sum_{i=1}^{n}(x_i^2) - (\sum_{i=1}^{n} x_i)^2} \\ &= \frac{\sum_{i=1}^{4} x_i^2 \sum_{i=1}^{4} y_i - \sum_{i=1}^{4} x_i \sum_{i=1}^{4}(x_i y_i)}{4\sum_{i=1}^{4}(x_i^2) - (\sum_{i=1}^{4} y_i)^2} \\ &= \frac{29 * 18 - 9 * 56}{4 * 29 - 81} \\ &= \frac{18}{35} \end{aligned}$$

(b) Substituting all the data points in equations in the form of y = mx+b:

$$\begin{aligned} 1 &= m0 + b \\ 3 &= m2 + b \\ 6 &= m3 + b \\ 8 &= m4 + b \end{aligned}$$

Re-writing into matrices in the form $X\theta = \mathbf{y}$:

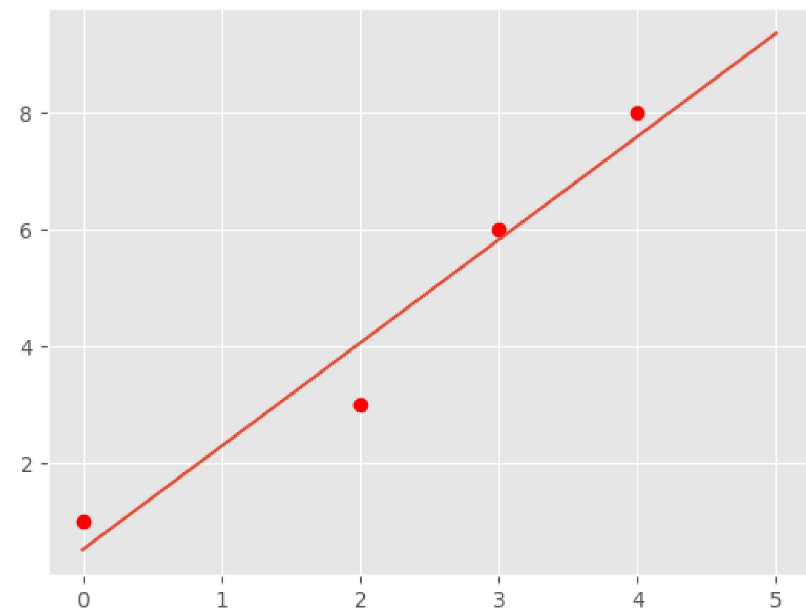
$$\begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$$

Using the normal equation:

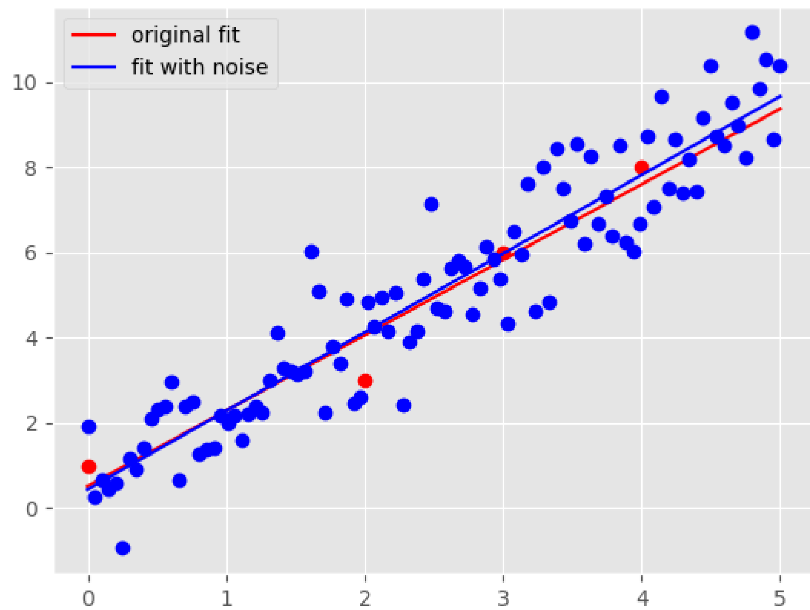
$$\begin{aligned}
\theta &= (X^T X)^{-1} X^T \mathbf{y} \\
&= (\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix})^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} \\
&= \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} \\
&= \frac{1}{35} \begin{bmatrix} 29 & -9 \\ -9 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} \\
&= \frac{1}{35} \begin{bmatrix} 29 & 11 & 2 & -7 \\ -9 & -1 & 3 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} \\
&= \frac{1}{35} \begin{bmatrix} 18 \\ 62 \end{bmatrix}
\end{aligned}$$

The answer is equal to that in part(a).

(c) Data with the optimal linear fit:




(d) Plot with randomly generated points with Gaussian noise:




The least squares estimate of the randomly generated points is
$y = 1.7634526262740544x + 0.6310995555484125$

$\blacksquare$