

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

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

(a) By definition, we know that

$$\mathbb{E}[\mathbf{y}] = \mathbb{E}[A\mathbf{x} + \mathbf{b}] = \int_{-\infty}^{+\infty} (A\mathbf{x} + \mathbf{b})\mathbb{P}(x)dx,$$

then, we can evaluate the integral and get:

$$\begin{aligned}\mathbb{E}[\mathbf{y}] &= \int_{-\infty}^{+\infty} (A\mathbf{x})\mathbb{P}(x)dx + \int_{-\infty}^{+\infty} (\mathbf{b})\mathbb{P}(x)dx \\ &= A \int_{-\infty}^{+\infty} \mathbf{x} \cdot \mathbb{P}(x)dx + \mathbf{b} \int_{-\infty}^{+\infty} \mathbb{P}(x)dx \\ &= A\mathbb{E}[\mathbf{x}] + \mathbf{b}(1) \\ &= A\mathbb{E}[\mathbf{x}] + \mathbf{b}\end{aligned}$$

(b) By definition, we know that

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

■

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

- (a) Find the least squares estimate $y = \theta^\top \mathbf{x}$ 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) let $y = \theta_0 + \theta_1 x$, and we know the design matrix to be:

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$$

From this, we can obtain:

$$X^\top X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}$$

Since we are looking for the values of θ_0 and θ_1 , and $X^\top X \theta = X^\top \mathbf{y}$, where $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$.

We can find the value of $X^\top \mathbf{y}$ at first and then use the Cramer's rule to find the value of θ_0 and θ_1 because $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$.

$$X^\top \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 18 \\ 56 \end{bmatrix}$$

Using Cramer's rule, we know

$$\theta_0 = \frac{\begin{vmatrix} 18 & 9 \\ 56 & 29 \end{vmatrix}}{\begin{vmatrix} 4 & 9 \\ 9 & 29 \end{vmatrix}} = \frac{18}{35}$$

and

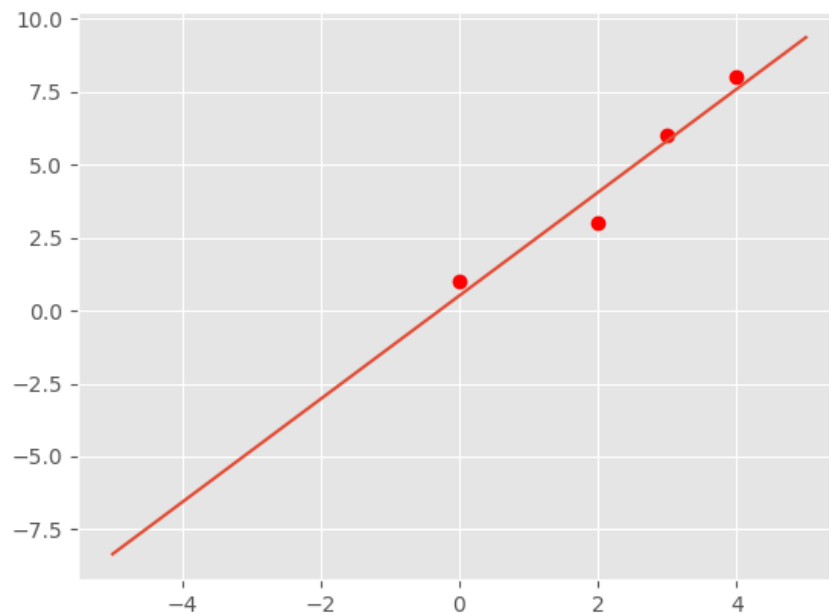
$$\theta_1 = \frac{\begin{vmatrix} 4 & 18 \\ 9 & 56 \end{vmatrix}}{\begin{vmatrix} 4 & 9 \\ 9 & 29 \end{vmatrix}} = \frac{62}{35}.$$

(b) Using normal equations, we can get:

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

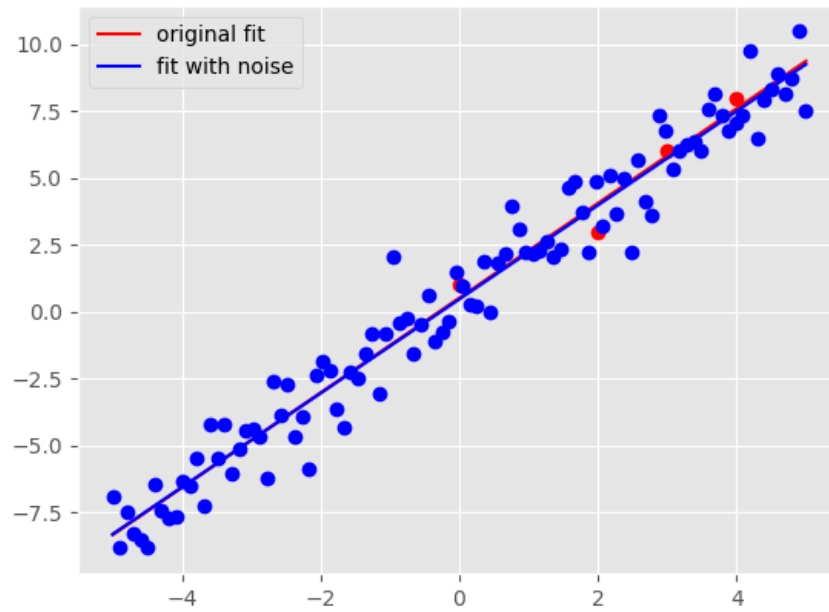
This confirms the value of θ_0 and θ_1 in part (a).

(c) Below is the fitted line based on the calculation of part (a) and part (b).



data with the calculated fit

(d) Below is the fitted line for 100 randomly generated points.



data with the calculated fit

As suggested by the plot, the fit with noise is incredibly close to the original fit. ■