Jacinda Chen
Math189R SU21
Homework 1
Friday, June 18, 2021

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

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

Since the probability of each y_i is equivalent to its corresponding x_i, the probability density function of $\mathbf{y}$ is equivalent to the density function of $\mathbf{x}$. Let $f(\mathbf{x})$ be the probability density function. Then, we can express the expectation of $\mathbf{y}$ as:

$$\begin{aligned}
\mathbb{E}[\mathbf{y}] &= \int_{\mathbb{R}} \mathbf{y} f(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}} (A\mathbf{x} + \mathbf{b}) f(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}} A\mathbf{x} f(\mathbf{x}) d\mathbf{x} + \int_{\mathbb{R}} \mathbf{b} f(\mathbf{x}) d\mathbf{x} \\
&= A \int_{\mathbb{R}} \mathbf{x} f(\mathbf{x}) d\mathbf{x} + \mathbf{b} \int_{\mathbb{R}} f(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

Since the total probability is 1, $\int_{\mathbb{R}} f(\mathbf{x}) d\mathbf{x} = 1$. Therefore,

$$\begin{aligned}
\mathbb{E}[\mathbf{y}] &= A \int_{\mathbb{R}} \mathbf{x} f(\mathbf{x}) d\mathbf{x} + b \\
&= A\mathbb{E}[\mathbf{x}] + \mathbf{b},
\end{aligned}$$

as desired. ■

Using the definition for covariance from lecture,

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

Since we have now proven that expectation is linear,

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

as desired. ∎

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) Given $\mathcal{D}$, we get the following sums

$$\sum_{i=1}^{4} x_i = 9 \qquad \sum_{i=1}^{4} y_i = 18 \qquad \sum_{i=1}^{4} x_i y_i = 56 \qquad \sum_{i=1}^{4} x_i^2 = 29.$$

Therefore, using Cramer's Rule, we get

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

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

Thus, the least squares estimate of this dataset is

$$\mathbf{y} = \left(\frac{18}{35}, \frac{62}{35}\right)^\top \mathbf{x}.$$

(b) For this dataset,

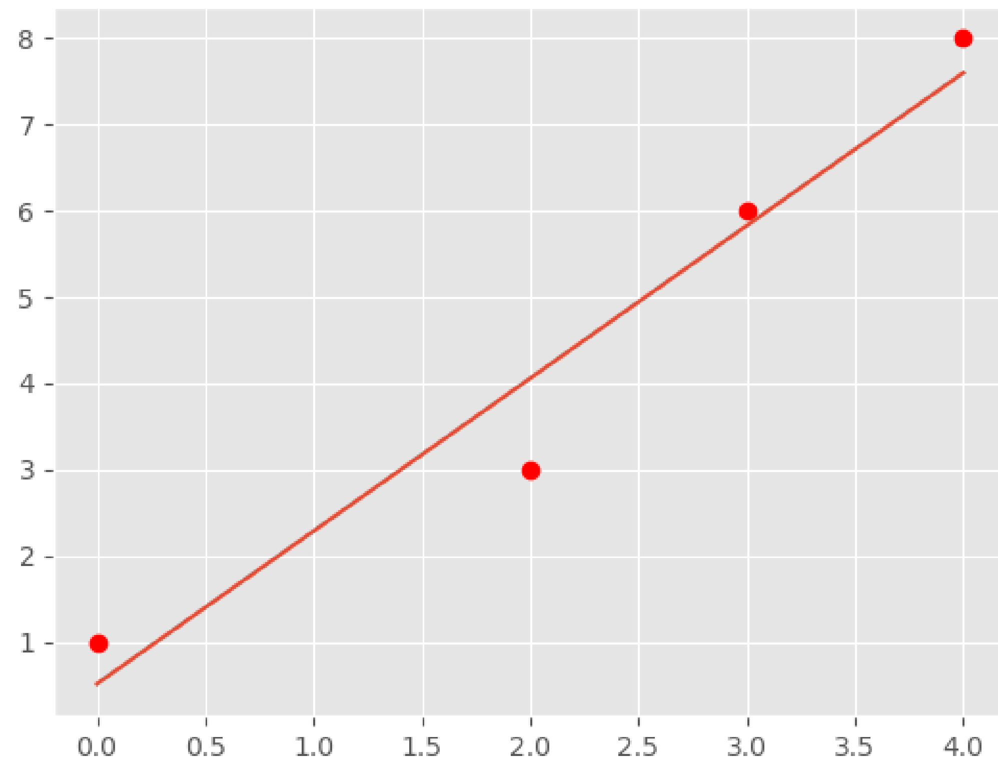
$$X = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}.$$

Therefore,

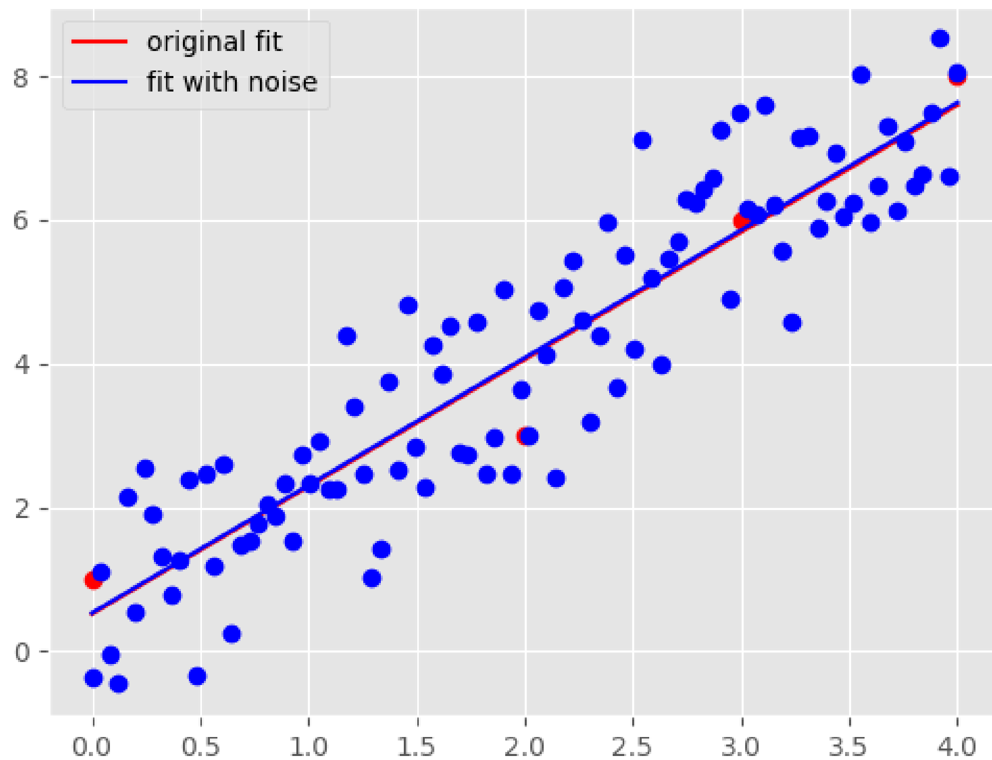
$$\begin{aligned}
\boldsymbol{\theta} &= (X^\top X)^{-1} X^\top \mathbf{y} \\
&= \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} \\
&= \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}^{-1} \begin{bmatrix} 18 \\ 56 \end{bmatrix} \\
&= \frac{1}{35} \begin{bmatrix} 29 & -9 \\ -9 & 4 \end{bmatrix} \begin{bmatrix} 18 \\ 56 \end{bmatrix} \\
&= \frac{1}{35} \begin{bmatrix} 18 \\ 62 \end{bmatrix} \\
&= \begin{bmatrix} 18/35 \\ 62/35 \end{bmatrix},
\end{aligned}$$

which matches our answer for part (a), as desired.

(c) The plot for this dataset is:



(d) The plot for the randomly generated 100 points is:



As seen in the plot, the original fit and the fit with noise are very similar.