Math189R SP19
Homework 1
Monday, Feb 4, 2017

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

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

We wish to show that $\mathbb{E}$[**y**] = $\mathbb{E}$[A**x**+ **b**] = A$\mathbb{E}$[**x**] + **b**. By definition of expectation, we have

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

where S is the space contains all possible x. Then, we can rearrange to get

$$\mathbb{E}[\mathbf{y}] = A\int_S \mathbf{x}\mathbb{P}(x)dx + \mathbf{b}\int_S \mathbb{P}(x)dx.$$

Note that by definition of expectation, $\mathbb{E}[\mathbf{x}] = \int_S \mathbf{x}\mathbb{P}(x)dx$. Also by definition of random variables $\int_S \mathbb{P}(x)dx = 1$. Thus, we can see that $\mathbb{E}[\mathbf{y}] = A\mathbb{E}[\mathbf{x}] + \mathbf{b}$.

Next, we wish to show that $\text{cov}[\mathbf{y}] = \text{cov}[A\mathbf{x} + \mathbf{b}] = A\text{cov}[\mathbf{x}]A^\top = A\mathbf{\Sigma}A^\top$. We are given $\mathbf{y} = A\mathbf{x} + \mathbf{b}$ so that $\text{cov}[\mathbf{y}] = \text{cov}[A\mathbf{x} + \mathbf{b}]$. Then, by definition of covariance, we have $\text{cov}[A\mathbf{x} + \mathbf{b}] = \mathbb{E}[(A\mathbf{x} + b - \mathbb{E}[A\mathbf{x} + b])(A\mathbf{x} + b - \mathbb{E}[A\mathbf{x} + b])^T]$. By linearity of expectation, we can rewrite as $\text{cov}[A\mathbf{x} + \mathbf{b}] = \mathbb{E}[(A\mathbf{x} + b - A\mathbb{E}[\mathbf{x}] - b)(A\mathbf{x} + b - A\mathbb{E}[\mathbf{x}] - b)^T]$. Simplifying, we get $\text{cov}[A\mathbf{x} + \mathbf{b}] = \mathbb{E}[(A\mathbf{x} - A\mathbb{E}[\mathbf{x}])(A(\mathbf{x} - \mathbb{E}[\mathbf{x}]))^T]$. Applying transpose in the second term inside expectation, we get $\text{cov}[A\mathbf{x} + \mathbf{b}] = \mathbb{E}[(A\mathbf{x} - A\mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T A^T]$. Since A and A^T are constants, by linearity, we have $\text{cov}[A\mathbf{x} + \mathbf{b}] = A\mathbb{E}[(\mathbf{x} - A\mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T]A^T$. Lastly, by definition of covariance, we see $\text{cov}[A\mathbf{x} + \mathbf{b}] = A\text{cov}[\mathbf{x}]A^T = A\mathbf{\Sigma}A^T$ 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) Let

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

To find the least square estimate using Cramer's Rule, we need to set $X^T X\theta = X^T \mathbf{y}$ where $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$. Plugging in, we get

$$X^T X\theta = \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = X^T \mathbf{y} = \begin{bmatrix} 18 \\ 56 \end{bmatrix}.$$

Next, we apply Cramer's Rule to solve for θ_0 and θ_1.

$$\theta_0 = \frac{\begin{vmatrix} 18 & 9 \\ 56 & 29 \end{vmatrix}}{\begin{vmatrix} 4 & 9 \\ 9 & 29 \end{vmatrix}} = \frac{18}{35} \text{ and } \theta_1 = \frac{\begin{vmatrix} 4 & 18 \\ 9 & 56 \end{vmatrix}}{\begin{vmatrix} 4 & 9 \\ 9 & 29 \end{vmatrix}} = \frac{62}{35}$$

Thus, for the given dataset, we found the least square estimate is $y = \frac{18}{35} + \frac{62}{35}x$ using Cramer's Rule

(b) Using the normal equation we can solve for θ using the following formula,

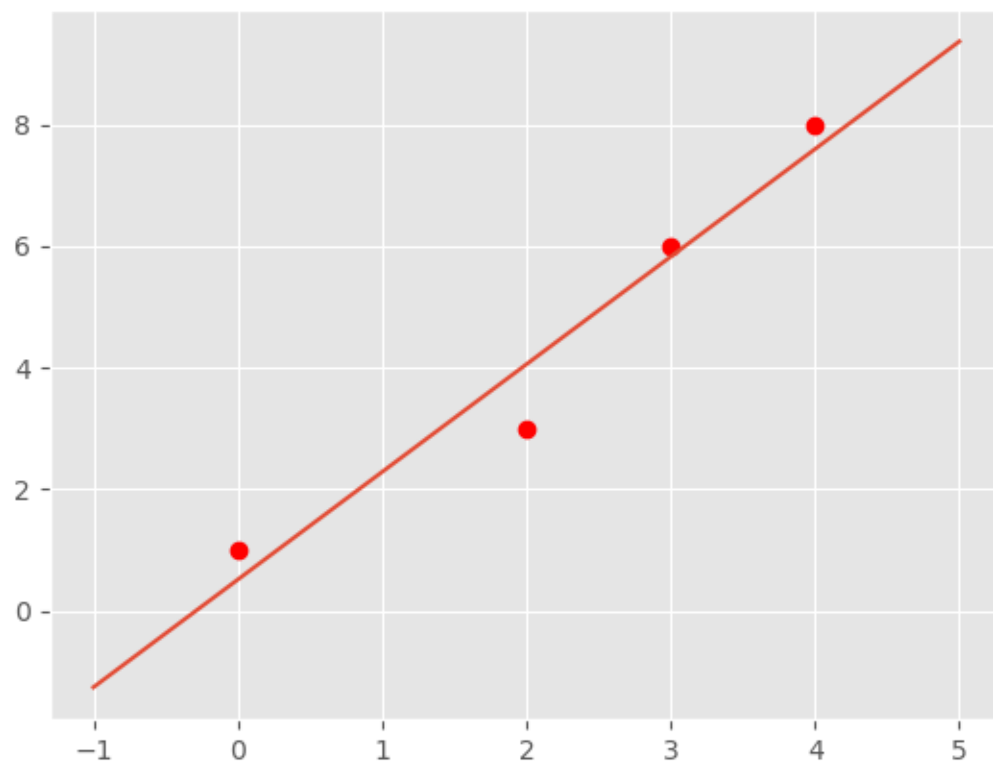
$$\theta = (X^T X)^{-1} X^T \mathbf{y}.$$

Then, we can plug in values from part a and evaluate,

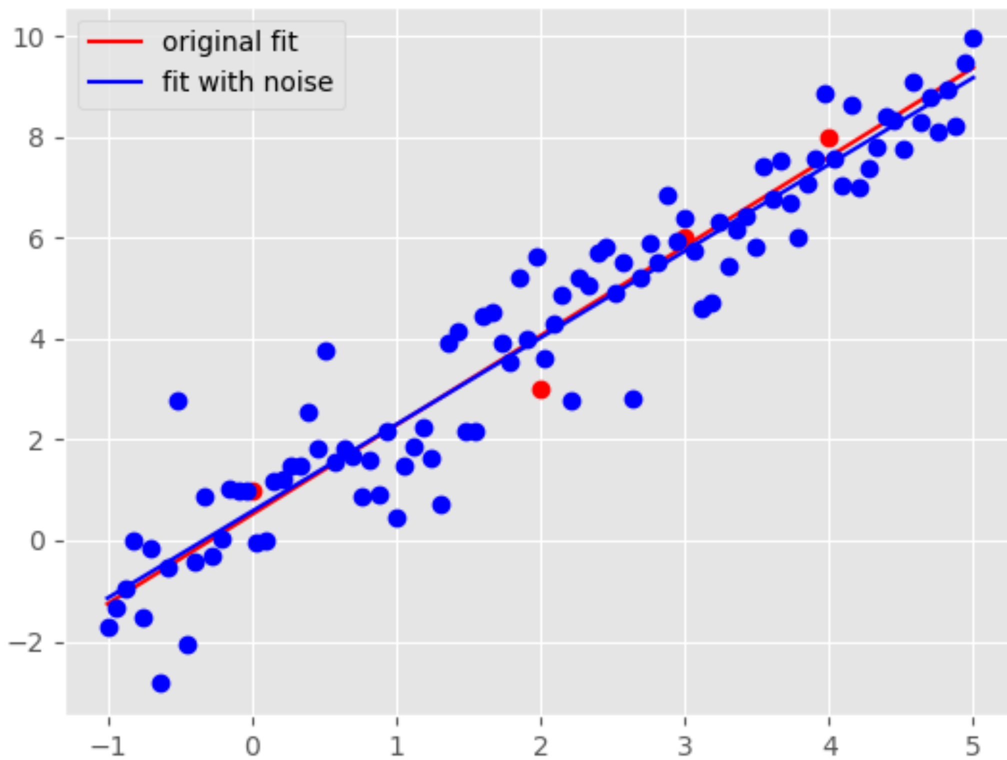
$$\theta = \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} \frac{18}{35} \\ \frac{62}{35} \end{bmatrix}.$$

Thus, we see that the solution using the normal equation is the same as using Cramer's rule as in part a.

(c) Plotted data and optimal linear fit.



(d) Plotted data with Gaussian noise and new least square estimate. Note that the new line is close to the old line.



∎