Colin Adams
Math189R SU20
Homework 1
Monday, June 2020

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.

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

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

For the first part, we have

$$\mathbb{E}[\mathbf{y}] = \mathbb{E}[A\mathbf{x} + \mathbf{b}] \equiv \int (A\mathbf{x} + \mathbf{b})\, p(\mathbf{x})\, \mathrm{d}\mathbf{x} = A\int \mathbf{x}\,\mathrm{d}\mathbf{x} + \int \mathbf{b}\, p(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Using the definition of $\mathbb{E}[\mathbf{x}] \equiv \int \mathbf{x}\,\mathrm{d}\mathbf{x}$ and since A and $\mathbf{b}$ are independent of $\mathbf{x}$, and—lastly—because $\int p(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$, we have

$$\mathbb{E}[\mathbf{y}] = \mathbb{E}[A\mathbf{x} + \mathbf{b}] = A\mathbb{E}[\mathbf{x}] + \mathbf{b}$$

as desired.

As for the second part, we have a similar strategy. Since the covariance is defined to be $\mathrm{cov}[\mathbf{x}] \equiv \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T]$, we have

$$\begin{aligned}
\mathrm{cov}[\mathbf{y}] = \mathrm{cov}[A\mathbf{x} + \mathbf{b}] &\equiv \mathbb{E}[(A\mathbf{x} + \mathbf{b} - \mathbb{E}[A\mathbf{x} + \mathbf{b}])(A\mathbf{x} + \mathbf{b} - \mathbb{E}[A\mathbf{x} + \mathbf{b}])^T] \\
&= \mathbb{E}[(A\mathbf{x} + \mathbf{b} - A\mathbb{E}[\mathbf{x}] - \mathbf{b})(A\mathbf{x} + \mathbf{b} - A\mathbb{E}[\mathbf{x}] - \mathbf{b})^T] && \text{(result from above)} \\
&= \mathbb{E}[(A\mathbf{x} - A\mathbb{E}[\mathbf{x}])(A\mathbf{x} - A\mathbb{E}[\mathbf{x}])^T] \\
&= \mathbb{E}[(A(\mathbf{x} - \mathbb{E}[\mathbf{x}]))(A(\mathbf{x} - \mathbb{E}[\mathbf{x}]))^T] \\
&= A\mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}]))(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T]A^T \\
&= A\mathrm{cov}[\mathbf{x}]A^T && \text{(def. of covariance)} \\
&= A\mathbf{\Sigma}A^T
\end{aligned}$$

as desired. ∎

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

(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.

This seems like it's going to be a bit tedious.

(a) We are hoping to solve for $y = mx + b$ for the least squares estimate using our data. There is a messy formula which I guess we will solve for. Notice the length of $\mathscr{D} = 4$. Also note the following

$$\sum_i x_i = 9, \qquad \sum_i x_i^2 = 29, \qquad \sum_i y_i = 18, \qquad \sum_i y_i^2 = 110, \qquad \sum_i x_i y_i = 56$$

which we plug into this hideous beasts of

$$m = \frac{n\sum_i x_i y_i - \left(\sum_i x_i\right)\left(\sum_i y_i\right)}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2} \quad \text{and} \quad b = \frac{\left(\sum_i x_i^2\right)\left(\sum_i y_i\right) - \left(\sum_i x_i\right)\left(\sum_i x_i y_i\right)}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}$$

which we evaluate and get

$$m = \frac{62}{35} \quad \text{and} \quad b = \frac{18}{35}$$

which gives us our old friend $y = mx + b$ as a fit for the data.

(b) Now, we use the normal equations to find this solution. We write

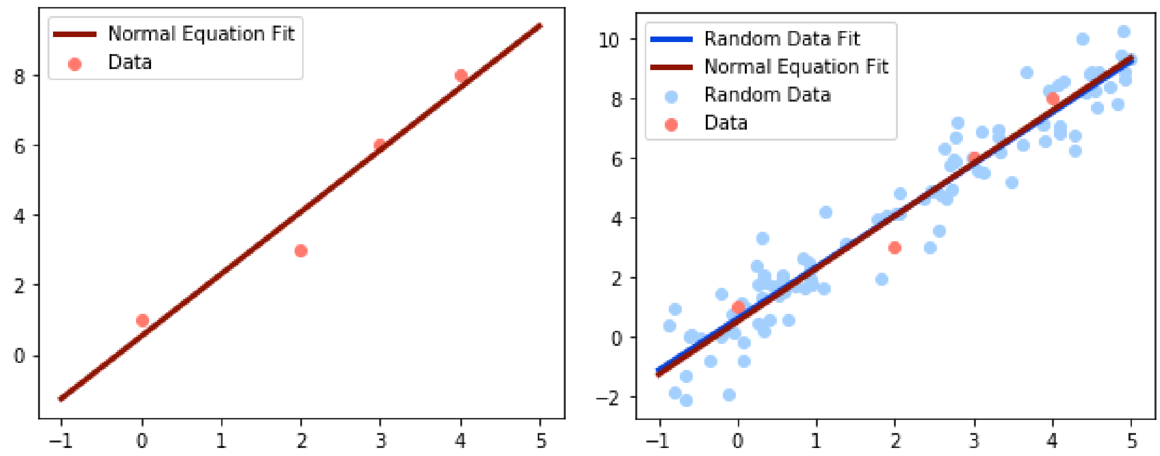
$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\theta} = \begin{bmatrix} b \\ m \end{bmatrix}.$$

Using these allows us to evaluate the normal equations

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

which is the result we expected.

(c)&(d) Now we will make the plot of our optimal linear fit as well as the one with the random noise.



These agree quite nicely with one another.

■