Cassidy Lê
Math189R SUP18
Homework 1
Tuesday, May 15, 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 $\mathbf{y} = A\mathbf{x} + b$ be a random vector. Show that expectation is linear:

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

Also show that

$$Cov[y] = Cov[Ax + b] = ACov[x]A^T = A\Sigma A^T.$$

(a) The expected value is defined as $\mathbb{E}[y] = \Sigma_{i=1}^{n} y_i p_i$. Given $\vec{y} = A\vec{x} + b$, substitute this into the summation for an expected value:

$$\mathbb{E}[y] = \Sigma_{i=1}^{n}(Ax_i + b)p_i.$$

Since $\vec{y} = A\vec{x} + b$, $p_i = \frac{1}{n}$ for all $i \in \mathbb{Z}$, so

$$\mathbb{E}[y] = \frac{1}{n}(Ax_1 + b) + \frac{1}{n}(Ax_2 + b) + ... + \frac{1}{n}(Ax_n + b)$$

$$= \frac{1}{n}A(x_1 + x_2 + ... + x_n) + \frac{1}{n}(nb)$$

$$= \frac{1}{n}A\vec{x} + b.$$

Since $p_i = \frac{1}{n}$ for all $i \in \mathbb{Z}$, $\mathbb{E}[x] = \frac{1}{n}(x_1 + x_2 + x_3 + ... + x_n) = \frac{1}{n}\vec{x}$. Therefore,

$$\mathbb{E}[y] = A\mathbb{E}[x] + b._{QED}$$

(b) Expanding $Cov[y]$ based on its definition results in

$$Cov[\vec{y}] = \mathbb{E}[(y_1 - \mathbb{E}[y_1])(y_2 - \mathbb{E}[y_2])...(y_n - \mathbb{E}[y_n])].$$

Since $\vec{y} = A\vec{x} + b$, plugging this into the covariance, we get

$$= \mathbb{E}[(Ax_1 + b - \mathbb{E}[Ax_1 + b])(Ax_2 + b - \mathbb{E}[Ax_2 + b])...(Ax_n + b - \mathbb{E}[Ax_n + b])].$$

From **part (a)**, we know that $\mathbb{E}[A\vec{x} + b] = A\mathbb{E}[\vec{x}] + b$. Thus,

$$= \mathbb{E}[(Ax_1 + b - A\mathbb{E}[x_1] - b)(Ax_2 + b - A\mathbb{E}[x_2] - b)...(Ax_n + b - A\mathbb{E}[x_n] - b)]$$

$$= \mathbb{E}[(Ax_1 - A\mathbb{E}[x_1])(Ax_2 - A\mathbb{E}[x_2])...(Ax_n - A\mathbb{E}[x_n])].$$

Because covariance matrices are symmetric and A is a constant,

$$= A\mathbb{E}[(x_1 - \mathbb{E}[x_1])(x_2 - \mathbb{E}[x_2])...(x_n - \mathbb{E}[x_n])]A^T$$

$$= ACov[\vec{x}]A^T$$

$$Cov[\vec{y}] = A\Sigma A^T ._{QED}$$

■

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

(a) Find the least squares estimate $y = \theta^T x$ 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) According to Cramer's Rule, we can use the following formulas to determine the slope m and y-intercept b for the least squares estimate:

$$m = \frac{n\Sigma_{i=1}^n x_i y_i - \left(\Sigma_{i=1}^n x_i\right)\left(\Sigma_{i=1}^n y_i\right)}{n\Sigma_{i=1}^n x_i^2 - \left(\Sigma_{i=1}^n x_i\right)^2}$$

$$b = \frac{\left(\Sigma_{i=1}^n x_i^2\right)\left(\Sigma_{i=1}^n y_i\right) - \left(\Sigma_{i=1}^n x_i\right)\left(\Sigma_{i=1}^n x_i y_i\right)}{n\Sigma_{i=1}^n x_i^2 - \left(\Sigma_{i=1}^n x_i\right)^2}.$$

First, we must determine the components:

$$\Sigma_{i=1}^4 x_i = 0+2+3+4 = 9$$

$$\Sigma_{i=1}^4 x_i^2 = 0^2+2^2+3^2+4^2 = 29$$

$$\Sigma_{i=1}^4 y_i = 1+3+6+8 = 18$$

$$\Sigma_{i=1}^4 x_i y_i = (0)(1)+(2)(3)+(3)(6)+(4)(8) = 56.$$

Plugging these values into the equations for m and b, we get

$$m = \frac{(4)(56)-(9)(18)}{(4)(29)-9^2} = \frac{62}{35}$$

$$b = \frac{(29)(18)-(9)(56)}{(4)(29)-9^2} = \frac{18}{35}.$$

Therefore, $\boxed{\theta^T = \begin{bmatrix} \frac{18}{35} & \frac{62}{35} \end{bmatrix} \text{ or } y = \frac{62}{35}x + \frac{18}{35}.}$

(b) The normal equation states that $\theta = (X^T X)^{-1} X^T \vec{y}$. From the given data points, we know that

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

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

With this, we can determine θ by first calculated $X^T X$:

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

Then, calculate $(X^T X)^{-1}$:

$$(X^T X)^{-1} = \frac{1}{(4)(29) - 9^2} \begin{bmatrix} 29 & -9 \\ -9 & 4 \end{bmatrix} = \frac{1}{35} \begin{bmatrix} 29 & -9 \\ -9 & 4 \end{bmatrix}.$$

Finally, we can determine θ:

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

$$\theta = \frac{1}{35} \begin{bmatrix} 29 & 11 & 2 & -7 \\ -9 & -1 & 3 & 7 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$$

$$\theta = \frac{1}{35} \begin{bmatrix} 18 \\ 62 \end{bmatrix}.$$

Thus, $\boxed{\theta^T = \begin{bmatrix} \frac{18}{35} & \frac{62}{35} \end{bmatrix} \text{ or } y = \frac{62}{35}x + \frac{18}{35}.}$

∎

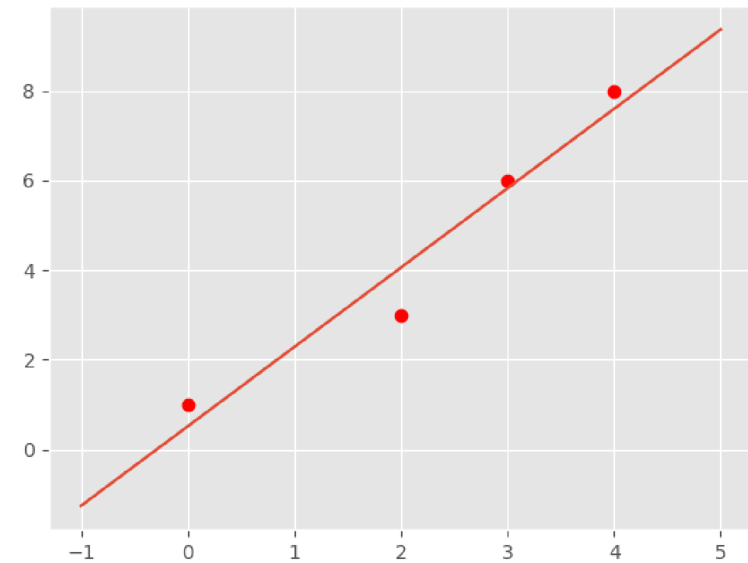

Figure 1: Optimal Linear Fit

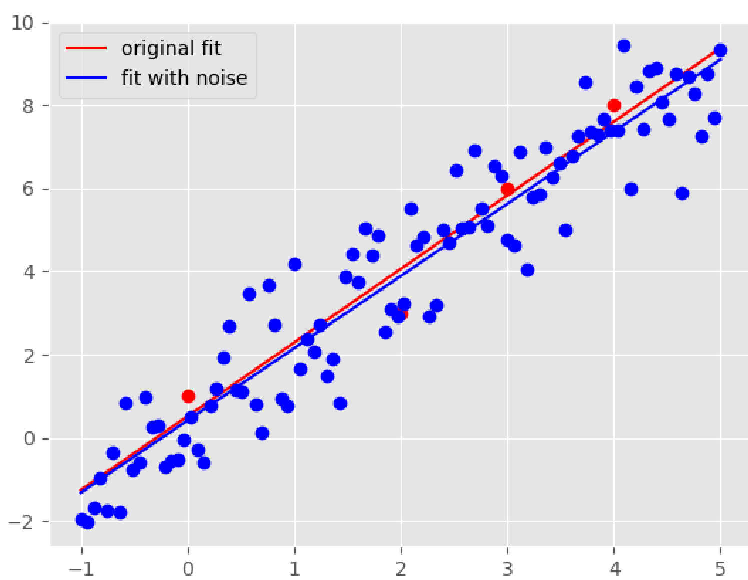

Figure 2: Optimal Linear Fit with Gaussian Noise