Math189R SU17
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 $\mathbf{y} = A\mathbf{x} + \mathbf{b}$ 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

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

We know that $E[u] = \int_U u \cdot p(u) \cdot du$ for some u. From this, we know that:

$$\mathbb{E}[\mathbf{y}] = \mathbb{E}[A\mathbf{x} + \mathbf{b}] = \int_X (A\mathbf{x} + \mathbf{b}) \cdot p(x) \cdot d\mathbf{x}$$

We may split up the terms of the integral to get:

$$\mathbb{E}[\mathbf{y}] = \int_X A\mathbf{x} \cdot p(\mathbf{x}) \cdot dx + \int_X \mathbf{b} \cdot p(\mathbf{x}) \cdot d\mathbf{x} = A\int_X \mathbf{x} \cdot p(\mathbf{x}) \cdot dx + \mathbf{b}\int_X p(\mathbf{x})d\mathbf{x}$$

Note that $\int_X \mathbf{x} \cdot p(\mathbf{x}) \cdot d\mathbf{x} = \mathbb{E}[\mathbf{x}]$, and that $\int_X p(\mathbf{x})d\mathbf{x} = 1$. Therefore,

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

Now we will show that

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

Recall that:

$$\mathrm{cov}[y] \triangleq \mathbb{E}[(\mathbf{y} - \mathbb{E}[\mathbf{y}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^T]$$

Substituting $\mathbf{y} = A\mathbf{x} + \mathbf{b}$ into this definition yields:

$$\mathrm{cov}[y] = \mathbb{E}[(A\mathbf{x} + \mathbf{b} - \mathbb{E}[A\mathbf{x} + b])(A\mathbf{x} + \mathbf{b} - \mathbb{E}[A\mathbf{x} + \mathbf{b}])^T]$$

We know that $\mathbb{E}[A\mathbf{x} + \mathbf{b}] = A\mathbb{E}[\mathbf{x}] + \mathbf{b}$. Therefore,

$$\mathrm{cov}[\mathbf{y}] = \mathbb{E}[(A\mathbf{x} - A\mathbb{E}[\mathbf{x}])(A\mathbf{x} - A\mathbb{E}[\mathbf{x}])^T] = A\mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T]A^T = A\mathrm{cov}[\mathbf{x}]A^T$$

Since $\mathbf{x}$ is a random vector of values, by definition,

$$A\mathrm{cov}[\mathbf{x}]A^T = A\Sigma A^T$$

■

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) We showed in class that the system of equations that minimises the loss function for parametres m and b for a linear model of N samples is given by:

$$\Sigma_i^N x_i y_i = m\Sigma_i^N x_i^2 + b\Sigma_i^N x_i$$

$$\Sigma_i^N y_i = m\Sigma_i^N x_i + bN$$

From the dataset $\mathcal{D}$, $\Sigma_i^N x_i y_i = 56$, $\Sigma_i^N x_i^2 = 29, \Sigma_i^N x_1 = 9, \Sigma_i^N y_i = 18$, and $N = 4$. We can now simplify the above system to:

$$9m + 4b = 18$$

$$29m + 9b = 56$$

We can encode this system in matrices like so:

$$\begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 18 \\ 56 \end{bmatrix}$$

Accordingly,

$$\begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}^{-1} \begin{bmatrix} 18 \\ 56 \end{bmatrix} = \frac{1}{35} \begin{bmatrix} 29 & -9 \\ -9 & 4 \end{bmatrix} \begin{bmatrix} 18 \\ 56 \end{bmatrix} = \frac{1}{35} \begin{bmatrix} 18 \\ 62 \end{bmatrix} = \begin{bmatrix} 0.51 \\ 1.77 \end{bmatrix}$$

The least squares estimate for y then becomes:

$$y = \frac{1}{35} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 18 \\ 62 \end{bmatrix} = \begin{bmatrix} 0.51 \\ 4.06 \\ 5.83 \\ 7.60 \end{bmatrix}$$

(b) The form of the normal equation we will use is $\theta = (X^T X)^{-1} X^T y$. Given

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

, the normal equation yields :

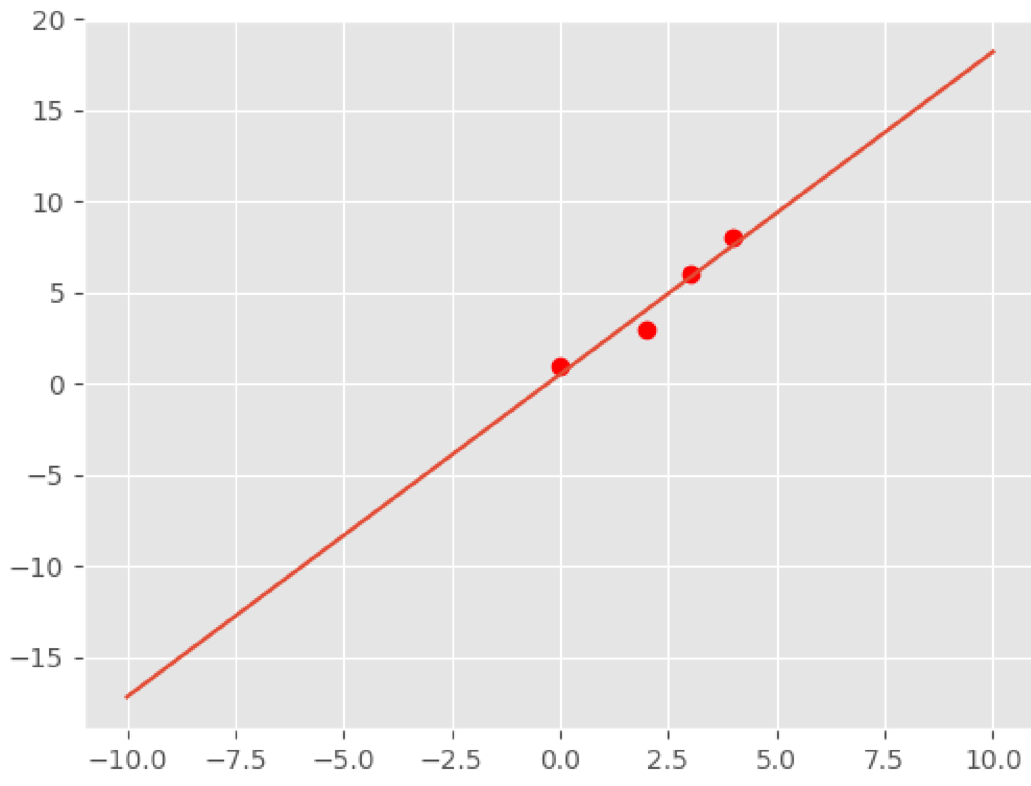
$$\theta = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$$

Evaluating this yields:

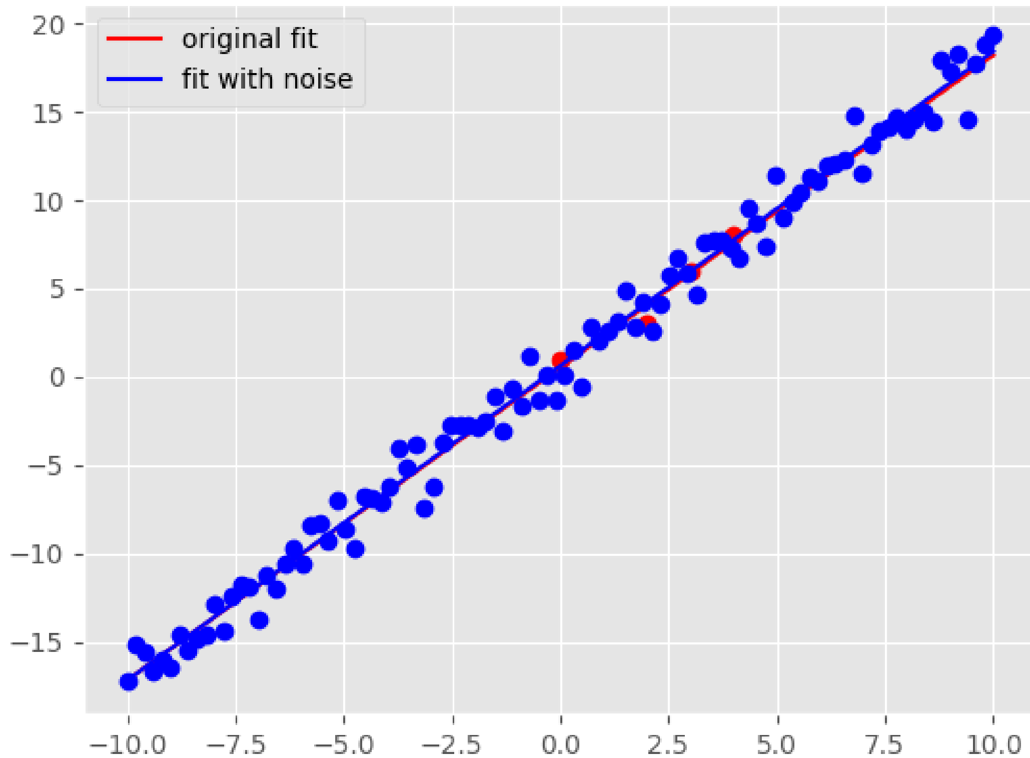
$$\theta = \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}^{-1} \begin{bmatrix} 18 \\ 56 \end{bmatrix} = \frac{1}{35} \begin{bmatrix} 18 \\ 62 \end{bmatrix} = \begin{bmatrix} 0.51 \\ 1.77 \end{bmatrix}$$

This is the same result we got in part (a), and will yield the estimate for y.

(c) The original data and the optimum linear fit I found with $m = 1.77$ and $b = 0.51$.



(d) A plot of the original fit and the fit with gaussian noise added.


original fit
fit with noise
20
15
10
5
0
−5
−10
−15
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0

■