Eli Weissler
Math189R SP19
Homework 1
Monday, February 03, 2019

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

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

First recall the definition of expectation value $\mathbb{E}$ of a vector $\mathbf{x}$

$$\mathbb{E}[\mathbf{x}] = \begin{pmatrix} \mathbb{E}[x_1] \\ \mathbb{E}[x_2] \\ \cdots \\ \mathbb{E}[x_i] \\ \cdots \end{pmatrix} = \begin{pmatrix} \int x_1 f(x_1) dx_1 \\ \int x_2 f(x_2) dx_2 \\ \cdots \\ \int x_i f(x_i) dx_i \\ \cdots \end{pmatrix}$$

where $f(x_i)$ is some probability density function, and the integral is performed over the kernel. In this case, we assume that $\mathbf{x}$ is a random vector with n continuous random variables, A is a constant matrix, and $\mathbf{b}$ is a constant vector.

$$\mathbb{E}[A\mathbf{x} + \mathbf{b}] = \begin{pmatrix} \mathbb{E}[A_{1:} \cdot \mathbf{x} + b_1] \\ \mathbb{E}[A_{2:} \cdot \mathbf{x} + b_2] \\ \cdots \\ \mathbb{E}[A_{i:} \cdot \mathbf{x} + b_i] \\ \cdots \end{pmatrix}$$

Note that I am using $A_{i:}$ to denote the row vector formed by all elements in row i of A. If we examine an arbitrary element, we see that

$$\mathbb{E}[A_{i:} \cdot \mathbf{x} + b_i] = \mathbb{E}[\sum_j A_{ij} x_k + b_i]$$

Now we use the formal definition of expectation value to convert to an integral. Although we have to integrate over all x_j, unless a term contains an x_j the integral goes to one.

$$\begin{aligned}
\mathbb{E}[\sum_j A_{ij}x_k + b_i] &= \int\int ... \int f(x_1)dx_1 f(x_2)dx_2 ... f(x_n)dx_n(\sum_j A_{ij}x_j f(x_j) + b_i) \\
&= \sum_j \int A_{ij}x_j f(x_j)dx_j + b_i = \sum_j A_{ij}\int x_j f(x_j)dx_j + b_i \\
&= \sum_j A_{ij}\mathbb{E}[x_j] + b_i = A_{i:}\cdot\mathbb{E}[\mathbf{x}] + b_i
\end{aligned}$$

Bringing this back together, we see that

$$\mathbb{E}[A\mathbf{x}+\mathbf{b}] = \begin{pmatrix} A_{1:}\cdot\mathbb{E}[\mathbf{x}] + b_1 \\ A_{2:}\cdot\mathbb{E}[\mathbf{x}] + b_2 \\ \cdots \\ A_{i:}\cdot\mathbb{E}[\mathbf{x}] + b_i \\ \cdots \end{pmatrix} = A\mathbb{E}[\mathbf{x}] + \mathbf{b}$$

as desired.

Now recall the definition of covariance

$$\text{cov}[\mathbf{x}]_{ij} = \mathbb{E}[(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])]$$

If we make the same assumptions, i.e. that $\mathbf{x}$ is a random vector with n continous variables, A is a constant matrix, and $\mathbf{b}$ is a constant vector, then

$$\begin{aligned}
cov[A\mathbf{x}+\mathbf{b}]_{ij} &= \mathbb{E}[((A\mathbf{x}+\mathbf{b})_i - \mathbb{E}[(A\mathbf{x}+\mathbf{b})_i])((A\mathbf{x}+\mathbf{b})_j - \mathbb{E}[(A\mathbf{x}+\mathbf{b})_j])] \\
&= \mathbb{E}[(A_{i:}\cdot\mathbf{x} + b_i - A_{i:}\cdot\mathbb{E}[x] - b_i)(A_{j:}\cdot\mathbf{x} + b_j - A_{j:}\cdot\mathbb{E}[x] - b_j)] \\
&= \mathbb{E}[(A_{i:}\cdot\mathbf{x} - A_{i:}\cdot\mathbb{E}[x]-)(A_{j:}\cdot\mathbf{x} - A_{j:}\cdot\mathbb{E}[x])] \\
&= \mathbb{E}[(A_{i:}\cdot\mathbf{x} - A_{i:}\cdot\mathbb{E}[x]-)(\mathbf{x}^T\cdot A_{:j}^T - \mathbb{E}[x]^T\cdot A_{:j}^T)] \quad \text{Rewriting the right parenthesis} \\
&= \mathbb{E}[A_{i:}\cdot(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x}^T - \mathbb{E}[\mathbf{x}]^T)\cdot A_{:j}] \\
&= A_{i:}\cdot\mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x}^T - \mathbb{E}[\mathbf{x}]^T)]\cdot A_{:j} \\
&= (A\text{cov}[\mathbf{x}]A^T)_{ij}
\end{aligned}$$

Therefore $\text{cov}[A\mathbf{x}+\mathbf{b}] = A\text{cov}[\mathbf{x}]A^\top$ as desired.

$\blacksquare$

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 are trying to minimize the cost function

$$J = \sum [y_i - (mx_i + b)]^2$$

We can do this by setting the partials with respect to m, b equal to zero.

$$\begin{aligned}
\frac{\partial J}{\partial m} &= \sum 2[y_i - (mx_i + b)](-x_i) = 0 \\
\frac{\partial J}{\partial b} &= \sum 2[y_i - (mx_i + b)](-1) = 0 \\
&\rightarrow \sum x_i y_i = m \sum x_i^2 + b \sum x_i \\
&\rightarrow \sum y_i = m \sum x_i + bN
\end{aligned}$$

In this problem, this is equivalent to the matrix

$$\begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix} = \begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & N \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix}$$
$$\begin{pmatrix} 56 \\ 18 \end{pmatrix} = \begin{pmatrix} 29 & 9 \\ 9 & 4 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix}$$

Using Cramer's rule from class, we get

$$m = \frac{62}{35}$$
$$b = \frac{18}{35}$$

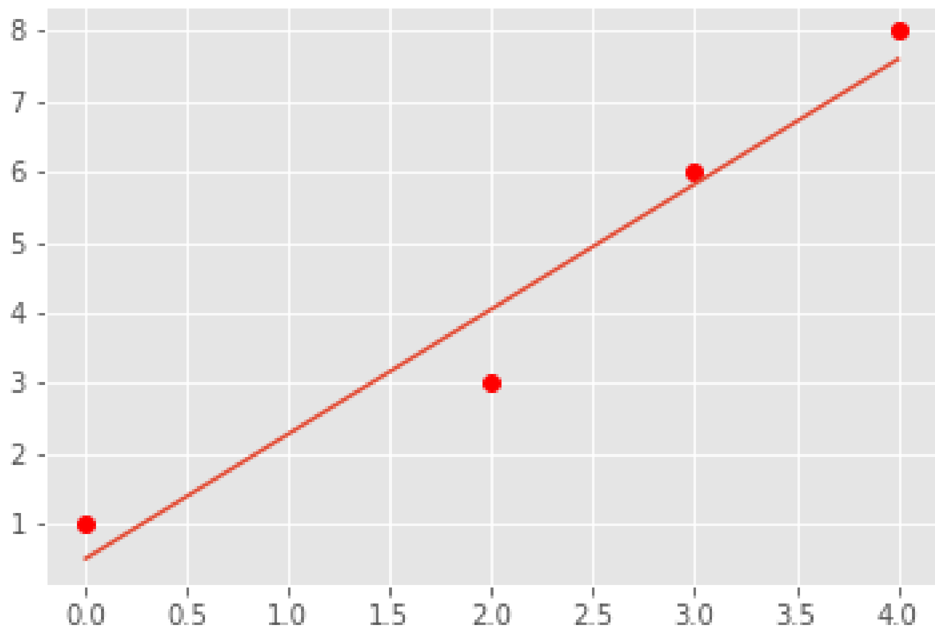
(b) Define our feature matrix and outcomes to be

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \qquad Y = \begin{pmatrix} 1 \\ 3 \\ 6 \\ 8 \end{pmatrix}$$

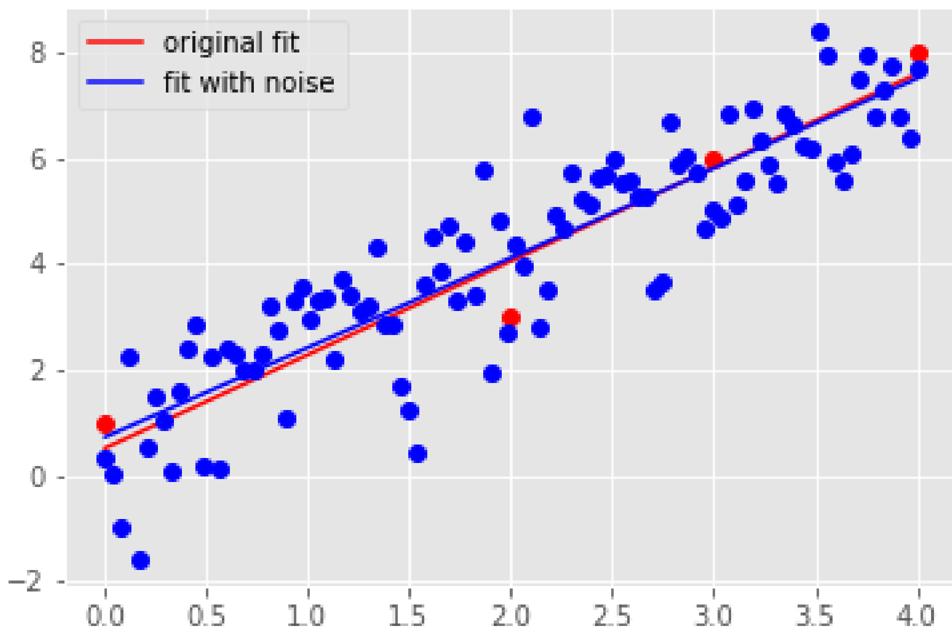
Which yields (I multiplied matrices using Mathematica)

$$\theta = (X^T X)^{-1}(X^T Y) = \begin{pmatrix} \frac{18}{35} \\ \frac{62}{35} \end{pmatrix}$$

Which agrees with the results from part (a), as $\theta_0 = b$ and $\theta_1 = m$.



(c)



(d)

■