Nico Espinosa Dice
Math189R SP19
Homework 1
Monday, Feb 4, 2017

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 will show linearity in the case where Y is not a vector, and we will then generalize to show linearity for $\mathbf{y}$:

$$\begin{aligned}
&\mathbb{E}[Y] = \mathbb{E}[AX + b] \\
&= \sum_{x \in X} (Ax + b)p(x) \text{ (by definition of expectation)} \\
&= \sum_{x \in X} Axp(x) + \sum_{x \in X} bp(x) \text{ (distributive property)} \\
&= \sum_{x \in X} Axp(x) + b \text{ (since } b \text{ is constant and the probabilities of } x_i \text{ sum to 1)} \\
&= A \sum_{x \in X} xp(x) + b \text{ (since A is constant)} \\
&= A\mathbb{E}[x] + b \text{ (by definition of expectation).}
\end{aligned}$$

Now we consider the case with $\mathbf{y}$:

$$
\begin{aligned}
\mathbb{E}[\mathbf{y}] &= \mathbb{E}[\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}] \\
&= \begin{bmatrix} \mathbb{E}[Y_1] \\ \vdots \\ \mathbb{E}[Y_n] \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{E}[AX_1 + b_1] \\ \vdots \\ \mathbb{E}[AX_n + b_n] \end{bmatrix} \\
&= \begin{bmatrix} A\mathbb{E}[X_1] + b_1 \\ \vdots \\ A\mathbb{E}[X_n] + b_n \end{bmatrix} \text{ (proved above)} \\
&= A \begin{bmatrix} \mathbb{E}[X_1] + b_1 \\ \vdots \\ \mathbb{E}[X_n] + b_n \end{bmatrix} \\
&= A \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix} + \mathbf{b} \\
&= A\mathbb{E}[\mathbf{x}] + \mathbf{b}.
\end{aligned}
$$

Thus, we have shown that expectation is linear. ■

Next, we will show that covariance is linear:

$$
\begin{aligned}
&\text{cov}[\mathbf{y}] = \text{cov}[A\mathbf{x} + \mathbf{b}] \\
&= \mathbb{E}[((A\mathbf{x} + \mathbf{b}) - \mathbb{E}[A\mathbf{x} + \mathbf{b}])((A\mathbf{x} + \mathbf{b}) - \mathbb{E}[A\mathbf{x} + \mathbf{b}])^\top] \text{ (definition of covariance)} \\
&= \mathbb{E}[(A\mathbf{x} + \mathbf{b} - A\mathbb{E}[\mathbf{x}] - \mathbf{b})(A\mathbf{x} + \mathbf{b} - A\mathbb{E}[\mathbf{x}] - \mathbf{b})^\top] \text{ (linearity of expectation)} \\
&= \mathbb{E}[(A\mathbf{x} - A\mathbb{E}[\mathbf{x}])(A\mathbf{x} - A\mathbb{E}[\mathbf{x}])^\top] \text{ (simplify algebra)} \\
&= \mathbb{E}[A(\mathbf{x} - \mathbb{E}[\mathbf{x}])((A\mathbf{x})^\top - (A\mathbb{E}[\mathbf{x}])^\top)] \text{ (property of transpose: sum/difference)} \\
&= \mathbb{E}[A(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x}^\top A^\top - \mathbb{E}[\mathbf{x}]^\top A^\top)] \text{ (property of tranpose: product)} \\
&= \mathbb{E}[A(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x}^\top - \mathbb{E}[\mathbf{x}]^\top)A^\top] \text{ (distributive property)} \\
&= \mathbb{E}[A(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top A^\top] \text{ (property of transpose: sum/difference)} \\
&= A\mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top]A^\top \text{ (linearity of expectation)} \\
&= A\text{cov}[\mathbf{x}]A^\top \text{ (definition of covariance)} \\
&= A \sum A^\top.
\end{aligned}
$$

■

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.

Github username: nico-espinosadice

(a) From Equation 7.15 in Murphy, we have:

$$\mathbf{X}^\top \mathbf{X} \boldsymbol{\theta} = \mathbf{X}^\top \mathbf{y}.$$

We define $\mathbf{X}$ as:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x_0} & \mathbf{x_1} \end{bmatrix}, \text{ where } \mathbf{x_0} = \mathbf{1} \text{ and } \mathbf{x_1} = \begin{bmatrix} 0 & 2 & 3 & 4 \end{bmatrix}.$$

Thus, we see that:

$$\begin{aligned}
&\mathbf{X}^\top \mathbf{X} \boldsymbol{\theta} = \mathbf{X}^\top \mathbf{y} \\
&\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} \\
&\begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 18 \\ 56 \end{bmatrix}.
\end{aligned}$$

By Cramer's Rule, we know that $b = \frac{D_x}{D}$ and $m = \frac{D_y}{D}$, where

$$\begin{aligned}
D &= \begin{vmatrix} 4 & 9 \\ 9 & 29 \end{vmatrix} = 35, \\
D_x &= \begin{vmatrix} 18 & 9 \\ 56 & 29 \end{vmatrix} = 18, \\
D_y &= \begin{vmatrix} 4 & 18 \\ 9 & 56 \end{vmatrix} = 62, \text{ so} \\
b &= \frac{18}{35}, \\
m &= \frac{62}{35}, \text{ and} \\
\boldsymbol{\theta} &= \begin{bmatrix} \frac{18}{35} \\ \frac{62}{35} \end{bmatrix}.
\end{aligned}$$

Thus,

$$\begin{aligned} y &= \boldsymbol{\theta}^\top \mathbf{x} \\ &= \begin{bmatrix} \frac{18}{35} & \frac{62}{35} \end{bmatrix} \mathbf{x}. \end{aligned}$$

∎

(b) The normal equation (Equation 7.16 in Murphy) is:

$$\begin{aligned} \boldsymbol{\theta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \\ &= \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} \\ &= \left(\begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix} \right)^{-1} \begin{bmatrix} 18 \\ 56 \end{bmatrix} \\ &= \begin{bmatrix} \frac{29}{35} & \frac{-9}{35} \\ \frac{-9}{35} & \frac{4}{35} \end{bmatrix} \begin{bmatrix} 18 \\ 56 \end{bmatrix} \\ &= \begin{bmatrix} \frac{18}{35} \\ \frac{62}{35} \end{bmatrix}. \end{aligned}$$

This solution found using the normal equation is the same as in part (a). ∎

(c)

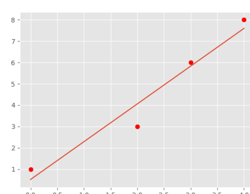

Figure 1: Plot of optimal linear fit.

(d)

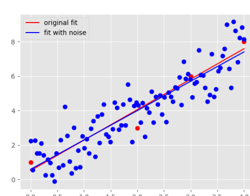

Figure 2: Plot of optimal linear fit, data with noise, and new fit for noisy data.