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

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

Part 1:
Since $\mathbb{E}$ is a weighted sum of all random variable, $\mathbb{E}$ is a linear operator. First prove that

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

for a random variable.
If X is a random variables, and a and b are constants,
Case 1: if X is discrete:

$$\begin{aligned}\mathbb{E}[aX + b] = &\sum_{x \in X}(ax + b)p(x) \\ =&a\sum_{x \in X} xp(x) + \sum_{x \in X} bp(x) \\ =&a\mathbb{E}[X] + b\end{aligned}$$

Case 2: if X is continuous:

$$\begin{aligned}\mathbb{E}[aX + b] = &\int_{\mathbb{R}}(ax + b)p(x)dx \\ =&a\int_{\mathbb{R}} xp(x)dx + b\int p(x)dx \\ =&a\mathbb{E}[X] + b\end{aligned}$$

Now, we show that this is true for a random vector: Let A be a constant $m \times n$ matrix, $\mathbf{x}$ be a length n random vetor, and $\mathbf{b}$ be a length m constant vector. Consider a single element in random vector $\mathbf{y}$:

$$\begin{aligned}\mathbb{E}[y_i] =&\mathbb{E}\left[\sum_{j=1}^{n} a_{i,j}x_j + b_i\right] \\ =&\sum_{j=1}^{n} a_{i,j}\mathbb{E}[x_j] + b_i\end{aligned}$$

Therefore we see that

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

$\blacksquare$

Part 2:
The prove this property, we first show that $\mathrm{cov}[\mathbf{z}+\mathbf{b}] = \mathrm{cov}[\mathbf{z}]$, then we show that $\mathrm{cov}[A\mathbf{x}] = A\mathrm{cov}[\mathbf{x}]A^T$.
For $\mathrm{cov}[\mathbf{z}+\mathbf{b}] = \mathrm{cov}[\mathbf{z}]$, let $\mathbf{z}$ be a length n random vector and $\mathbf{b}$ be a constant n-length vector. Let $\Sigma = \mathrm{cov}[\mathbf{z}+\mathbf{b}]$.

$$\begin{aligned}\Sigma_{i,j} =&\mathrm{cov}[z_i + b_i, z_j + b_j] \\ =&\mathbb{E}\left[((z_i + a_i) - \mathbb{E}[z_i + a_i])((z_j + a_j) - \mathbb{E}[z_j + a_j]))\right] \\ =&\mathbb{E}\left[(z_i + a_i - a_i - \mathbb{E}[z_i])(z_j + a_j - a_j - \mathbb{E}[z_j]))\right] \\ =&\mathbb{E}\left[(z_i - \mathbb{E}[x_i])(z_j - \mathbb{E}[z_j]))\right] \\ =&\mathrm{cov}[z_i + b_i, z_j + b_j]\end{aligned}$$

Therefore, we see that $\mathrm{cov}[\mathbf{z}+\mathbf{b}] = \mathrm{cov}[\mathbf{z}]$.
For $\mathrm{cov}[A\mathbf{x}] = A\mathrm{cov}[\mathbf{x}]A^T$, let A be a $m \times n$ matrix and let $\mathbf{x}$ be a length n vector. Let $\Sigma' = \mathrm{cov}[A\mathbf{x}]$. Consider element (i,j) in Σ'.

$$\begin{aligned}\Sigma'_{i,j} =&\mathrm{cov}\left(\sum_{k=1}^{n} a_{i,k}x_k, \sum_{k=1}^{n} a_{j,k}x_k\right] \\ =&\mathbb{E}\left[\left(\sum_{k=1}^{n} a_{i,k}x_k - \mathbb{E}\left[\sum_{k=1}^{n} a_{i,k}x_k\right]\right)\left(\sum_{k=1}^{n} a_{j,k}x_k - \mathbb{E}\left[\sum_{k=1}^{n} a_{j,k}x_k\right]\right)\right] \\ =&\mathbb{E}\left[\left(\sum_{k=1}^{n} a_{i,k}x_k - \sum_{k=1}^{n} a_{i,k}\mathbb{E}[x_k]\right)\left(\sum_{k=1}^{n} a_{j,k}x_k - \sum_{k=1}^{n} a_{j,k}\mathbb{E}[x_k]\right)\right] \\ =&\mathbb{E}\left[\sum_{k=1}^{n} a_{i,k}\left(x_k - \mathbb{E}[x_k]\right)\left(x_k - \mathbb{E}[x_k]\right)\sum_{k=1}^{n} a_{j,k}\right] \\ =&\sum_{k=1}^{n} a_{i,k}\left(\mathbb{E}\left[\left(x_k - \mathbb{E}[x_k]\right)\left(x_k - \mathbb{E}[x_k]\right)\right]\right)\sum_{k=1}^{n} a_{j,k}\end{aligned}$$

From this, we see that $\Sigma' = A\text{cov}[\mathbf{x}]A^T$, If we let $\text{cov}[\mathbf{x}] = \Sigma$, then $\Sigma' = A\Sigma A^T$.

Since we have proven both of these properties, we can simply substitute $\mathbf{z}$ for $A\mathbf{x}$:

$$\begin{aligned}\text{cov}[A\mathbf{x} + \mathbf{b}] &= \text{cov}[\mathbf{z} + \mathbf{b}] \\ &= \text{cov}[\mathbf{z}] \\ &= \text{cov}[A\mathbf{x}] \\ &= A\Sigma A^T\end{aligned}$$

$\blacksquare$

$\blacksquare$

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)

x_i	y_i	x_i^2	$x_i y_i$
0	1	0	0
2	3	4	6
3	6	9	18
4	8	16	32

$$
\begin{aligned}
m =& \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\left(n \sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \\
=& \frac{4 \cdot 56 - 9 \cdot 18}{4 \cdot 29 - 81} \\
\approx& 1.77 \\
b =& \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \\
=& \frac{29 \cdot 19 - 9 \cdot 56}{4 \cdot 29 - 81} \\
=& -6.94
\end{aligned}
$$

(b)

$$
\begin{aligned}
\theta =& (X^T X)^{-1} X^T \mathbf{y} \\
=& 29^{-1} \begin{bmatrix} 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} \\
\approx& 1.93
\end{aligned}
$$

■