Paul Steller - Homework 1
MATH 637

Solution 1.

We wish to prove that the problem will always have a unique solution, given by

$$\hat{\beta} = (X^T X + \alpha I)^{-1} X^T y$$

Similar to what we did in class, we can minimize $||y - X\beta||_2^2 + \alpha||\beta||_x^2$ by taking its derivative and setting it equal to 0.

$$\frac{\partial}{\partial \beta_i}(||y - X\beta||_2^2 + \alpha||\beta||_2^2) = 0$$

$$\frac{\partial}{\partial \beta_i} \sum_{k=1}^{n} (y_k - X_{k1}\beta_1 - \cdots - X_{kp}\beta_p)^2 + \alpha \sum_{k=1}^{p} (\beta_k)^2 = 0$$

$$2(X^T X\beta - X^T y) + 2\alpha\beta = 0$$

$$X^T X\beta + \alpha\beta = X^T y$$

$$(X^T X + \alpha I)\beta = X^T y$$

$$\beta = (X^T X + \alpha I)^{-1} X^T y$$

So we get our desired solution $\hat{\beta}$.
Next, we look at the Hessian:

$$\frac{\partial^2}{\partial \beta_i \partial \beta_j}(||y - X\beta||_2^2 + \alpha||\beta||_x^2) = \frac{\partial^2}{\partial \beta_i \partial \beta_j}(||y - X\beta||_2^2) + \frac{\partial^2}{\partial \beta_i \partial \beta_j}(\alpha||\beta||_2^2)$$

$$= 2X^T X + 2\alpha I$$

$X^T X$ is positive semi-definite, and since $\alpha > 0, 2\alpha I$ is positive definite, and so $2X^T X + 2\alpha I$ is positive definite, proving that $\hat{\beta}$ is unique.

Solution 2. (a) First, we wish to show that there is no $\beta \in \mathbb{R}^2$ such that $y = X\beta$. Let's assume, as a contradiction, that there exists such a β.

$$y = X\beta$$

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

From this, we get the following system of equations:

$$\begin{cases} \beta_1 + 2\beta_2 = 1 \\ 2\beta_1 + \beta_2 = 1 \\ 3\beta_1 + \beta_2 = 1 \end{cases}$$

Solving the first two equations, we find that $\beta_1 = \frac{1}{3}$ and $\beta_2 = \frac{1}{3}$. Substituting this into the third equation, we get

$$3(\frac{1}{3}) + \frac{1}{3} = \frac{4}{3} \neq \frac{1}{3}$$

Because $\beta_1 = \frac{1}{3}$ and $\beta_2 = \frac{1}{3}$ is the *unique* solution for the first two equations, we arrive at a contradiction, which implies that there is no β that will give an exact solution to $y = X\beta$.

(b) Now we wish to show that $X^T X\beta = X^T y$ has a unique solution $\hat{\beta}$.

$$X^T X\beta = X^T y$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 14 & 7 \\ 7 & 6 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

From this, we get the system of equations:

$$\begin{cases} 14\beta_1 + 7\beta_2 = 6 \\ 7\beta_1 + 6\beta_2 = 4 \end{cases}$$

Because this is a system of linear equations, it will either have one (unique) solution, an infinite number of solutions, or no solution. Solving this system, we get that $\beta_1 = \frac{8}{35}$ and $\beta_2 = \frac{14}{35}$. Thus, we have found the unique solution:

$$\hat{\beta} = \begin{pmatrix} \frac{8}{35} \\ \\ \frac{14}{35} \end{pmatrix}$$

(c) Finally, we wish to compute $||y - X\hat{\beta}||_2$. First we see that

$$y - X\hat{\beta} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 8/35 \\ 14/35 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 36/35 \\ 30/35 \\ 38/35 \end{pmatrix} = \begin{pmatrix} -1/35 \\ 5/35 \\ -3/35 \end{pmatrix}$$

Then,

$$|| \begin{pmatrix} -1/35 \\ 5/35 \\ -3/35 \end{pmatrix} ||_2 = \sqrt{(\frac{-1}{35})^2 + (\frac{5}{35})^2 + (\frac{-3}{35})^2} = \sqrt{\frac{1}{35^2} + \frac{25}{35^2} + \frac{9}{35^2}}$$

$$= \sqrt{\frac{35}{35^2}} = \sqrt{\frac{1}{35}} = \frac{1}{\sqrt{35}}$$

From class, we showed that the solution $\hat{\beta}$ of the normal equations $X^T X\beta = X^T y$ will minimize $||y - X\beta||_2^2$. Since our $\hat{\beta}$ was found this way, it minimizes $||y - X\beta||_2^2$, and thus

$$||y - X\beta||_2 \geq ||y - X\hat{\beta}||_2 \quad \forall \beta \in \mathbb{R}^2.$$

Solution 3. (a) Using the hint, we have that

$$X^T(X\hat{\beta} - y) = -X^T\hat{\epsilon} = 0$$

We don't know exactly what X^T looks like, but we do know that the first row is all ones. The rest of the matrix will be left blank to signify that it is unknown:

$$-X^T\hat{\epsilon} = 0$$

$$-\begin{bmatrix} 1 & 1 & \cdots & 1 \\ & & & \\ & & & \\ & & & \end{bmatrix}\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$-\begin{bmatrix} \epsilon_1 + \epsilon_2 + \cdots + \epsilon_n \\ \\ \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$-\begin{bmatrix} \sum_{i=1}^n \epsilon_i \\ \\ \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Thus,

$$\sum_{i=1}^n \epsilon_i = 0 \Rightarrow \frac{1}{n}\sum_{i=1}^n \epsilon_i = 0$$

and we are done.

(b) Finally, we wish to find a simple example to show that the residuals may not have mean zero if an intercept is not included in the model. I suspect that using a linear regression (without an intercept) to model a constant function will be problematic. Say that we are trying to find the best fit for the three points $(0, 5), (1, 5)$, and $(2, 5)$. Clearly, if we were allowed to use an intercept, the line $y = 5$ would perfectly model the data (and the sum of the residuals would be zero). With the restriction of no intercept, however, python finds the line of best fit to be $y = 3x$, which has mean residual equal to 2. The python code and graph of the data/line of best fit are provided on the attached sheet.

As another simple example, we look at our solution to Problem 2. Note that

$$X = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$$

Since a column of ones was not added, there is no y-intercept in this model. After finding $\hat{\beta}$, we computed $y - X\hat{\beta}$ and got

$$y - X\hat{\beta} = \begin{pmatrix} -1/35 \\ 5/35 \\ -3/35 \end{pmatrix}$$

These residuals clearly do not sum to zero, and thus do not have mean zero.

In [7]:

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
```

In [8]:

```
x_0 = [0,1,2]
y_0 = [5,5,5]
x = np.array(x_0)
y = np.array(y_0)
x = x.reshape(len(x),1)
y = y.reshape(len(y),1)
```

In [11]:

```
linreg = LinearRegression(fit_intercept = False)
```

In [12]:

```
linreg.fit(x,y)
print(linreg.coef_)
print(linreg.intercept_)
```

```
[[3.]]
0.0
```

In [13]:

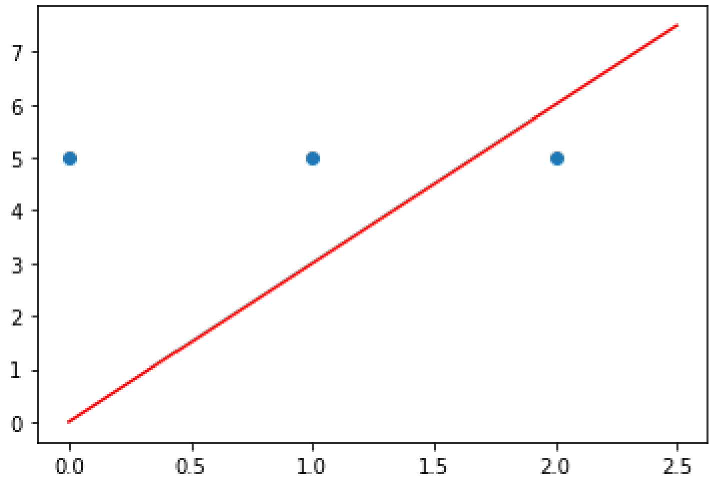
```
(y - linreg.predict(x)).mean()
```

Out[13]:

```
2.0
```

In [14]:

```
plt.scatter(x,y)
x_1 = np.linspace(0,2.5,100)
y_1 = 3 * x_1
plt.plot(x_1, y_1, c = 'red')
plt.show()
```



In [ ]:

```

```