Homework 0

Kaitlyn Mulligan

1/30/19

1 Python

Below is the script I used to install the packages.

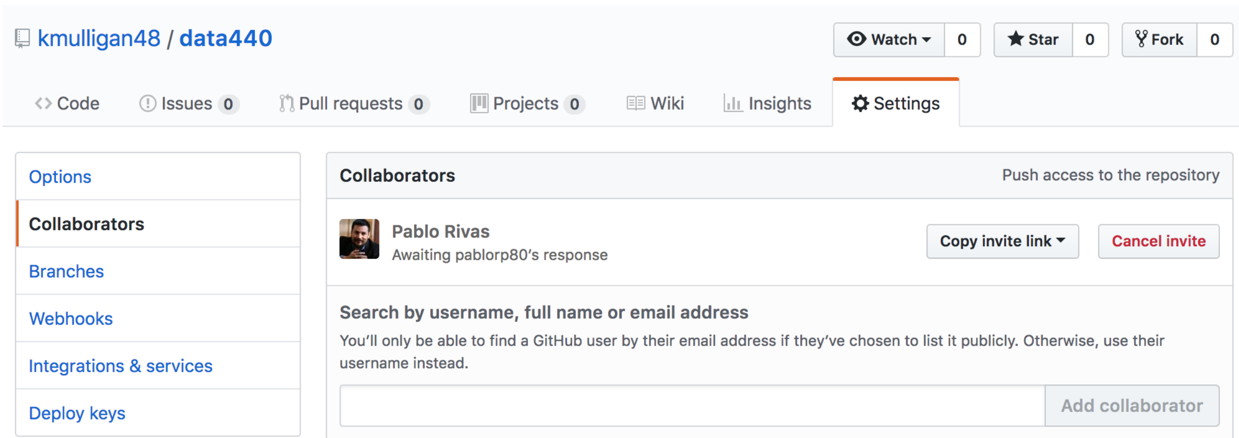
```
import sys
import numpy
import scipy
import sklearn
import matplotlib
import pandas
print(sys.version)
print(numpy.__version__)
print(scipy.__version__)
print(sklearn.__version__)
print(matplotlib.__version__)
print(pandas.__version__)
```

Here is the output of the print statements.

```
[148-100-154-221:mulligan-00 kaitlynmulligan$ python3 pythonInstall.py
3.7.2 (v3.7.2:9a3ffc0492, Dec 24 2018, 02:44:43)
[Clang 6.0 (clang-600.0.57)]
1.16.0
1.2.0
0.20.2
3.0.2
0.24.0
```

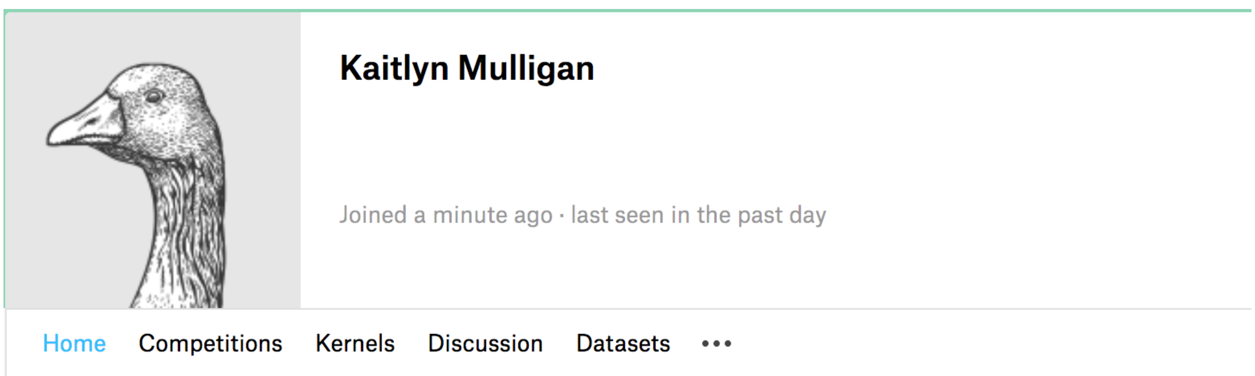
2 GitHub

My username on GitHub is **kmulligan48**. The link to my repository is **https://github.com/kmulligan48/data440.git**. Below is a screenshot to show that Professor Rivas has been added as a collaborator.



3 Kaggle

My Kaggle username is **kmulligan1** and below is a screenshot of my account.



4 Problems

1. For the function $g(x) = -3x^2 + 24x - 30$, find the value of x that maximizes $g(x)$.

Solution: First we take the derivative of $g(x)$ which results in

$$g'(x) = -6x + 24.$$

Next, set the derivative equal to zero and solve for x. This results in $x = 4$. To test to see if that is a maximum we can plug in a number less than and greater than $x = 4$ to the derivative. We can see that less than 4, the derivative is positive and greater than 4, the derivative is negative. Therefore, $x = 4$ is a value of x that maximizes $g(x)$.

2. Consider the following function:

$$f(x) = 3x_0^3 - 2x_0x_1^2 + 4x_1 - 8 \tag{1}$$

what are the partial derivatives of $f(x)$ with respect to x_0 and x_1.

Solution: The partial derivative of $f(x)$ with respect to x_0 is:

$$\frac{\partial f}{\partial x_0} = 9x_0^2 - 2x_1^2.$$

The partial derivative of $f(x)$ with respect to x_1 is:

$$\frac{\partial f}{\partial x_1} = -4x_0x_1 + 4.$$

3. Consider the matrix $A = \begin{bmatrix} 1 & 4 & -3 \\ 2 & -1 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} -2 & 0 & 5 \\ 0 & -1 & 4 \end{bmatrix}$, then answer the following and verify your answers in Python:
(a) can you multiply the two matrices? Ellaborate on your answer.

Solution: No, you cannot multiply these two matrices. In order to multiply two matrices, the number of columns of the first matrix needs to equal the number of rows of the second matrix. In this case, the number of columns in matrix A is 3, which does not equal the number of rows in matrix B which is 2.
Using Python to verify my answers, I ran the script below.

```
import numpy as np
A = np.array([[1,4,-3], [2,-1,3]])
B = np.array([[-2,0,5], [0,-1,4]])
A.dot(B)
```

This resulted in the following output which proves that we cannot multiply these two matrices.

```
Traceback (most recent call last):
  File "badMatrix.py", line 4, in <module>
    A.dot(B)
ValueError: shapes (2,3) and (2,3) not aligned: 3 (dim 1) != 2 (dim 0)
```

(b) multiply A^T and B and give its *rank*.

Solution: We begin by finding the transpose of A which is the following:

$$A^T = \begin{bmatrix} 1 & 2 \\ 4 & -1 \\ -3 & 3 \end{bmatrix}.$$

Next we can multiply A^T and B as follows:

$$A^TB = \begin{bmatrix} 1 & 2 \\ 4 & -1 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} -2 & 0 & 5 \\ 0 & -1 & 4 \end{bmatrix} = \begin{bmatrix} -2 & -2 & 13 \\ -8 & 1 & 16 \\ 6 & -3 & -3 \end{bmatrix}.$$

To find the rank, we need to reduce it to reduced row echelon form as follows:

$$\begin{bmatrix} -2 & -2 & 13 \\ -8 & 1 & 16 \\ 6 & -3 & -3 \end{bmatrix} \to \begin{bmatrix} -2 & -2 & 13 \\ 0 & 9 & -36 \\ 0 & -9 & 36 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & \frac{-13}{2} \\ 0 & 1 & -4 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & \frac{-5}{2} \\ 0 & 1 & -4 \\ 0 & 0 & 0 \end{bmatrix}.$$

As we can see, there are two non-zero rows. Therefore, the rank is 2.
Using Python to verify my answers, I ran the script below.

```
import numpy as np
A = np.array([[1,4,-3], [2,-1,3]])
B = np.array([[-2,0,5], [0,-1,4]])
print(A.T.dot(B))
```

This resulted in the following output which verifies the answers I got.

```
[[-2 -2 13]
 [-8  1 16]
 [ 6 -3 -3]]
```

(c) (extra credit) let $C = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ be a new matrix; what is the result of $AB^T + C^{-1}$?

Solution: We begin by finding the transpose of B which is the following:

$$B^T = \begin{bmatrix} -2 & 0 \\ 0 & -1 \\ 5 & 4 \end{bmatrix}.$$

Next we need to find the inverse of C as follows:

$$C^{-1} = \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

Now we can solve $AB^T + C^{-1}$ as follows:

$$AB^T + C^{-1} = \begin{bmatrix} 1 & 4 & -3 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 0 & -1 \\ 5 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -17 & -16 \\ 11 & 13 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -16 & -16 \\ 11 & \frac{27}{2} \end{bmatrix}.$$

Therefore, $AB^T + C^{-1} = \begin{bmatrix} -16 & -16 \\ 11 & \frac{27}{2} \end{bmatrix}$.

Using Python to verify my answers, I ran the script below.

```python
import numpy as np
A = np.array([[1,4,-3], [2,-1,3]])
B = np.array([[-2,0,5], [0,-1,4]])
C = np.array([[1,0], [0,2]])
print(A.dot(B.T)+np.linalg.inv(C))
```

This resulted in the following output which verifies the answers I got.

```
[[-16.  -16. ]
 [ 11.   13.5]]
```

4. Give the mathematical definitions of the simple Gaussian, multivariate Gaussian, Bernoulli, binomial, and exponential distributions.

Solution: **(a)** simple Gaussian distribution - otherwise known as the normal distribution, has a bell shaped curve, the pdf is given as $\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

(b) multivariate Gaussian distribution - otherwise known as multivariate normal distribution or joint normal distribution, it is a generalization of the one-dimensional normal distribution to a higher dimension, the pdf is given as $(2\pi)^{\frac{-k}{2}}|\Sigma|^{\frac{-1}{2}}e^{\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$

(c) Bernoulli distribution - a discrete distribution having two possible outcomes labelled by $x = 0$ or $x = 1$ in which $x = 1$ occurs with probability p and $x = 0$ occurs with probability $q = 1-p$, where $0 < p < 1$, the pdf is given as $p^x(1-p)^{1-x}$

(d) binomial distribution - the discrete probability distribution, with parameters n and p, of the number of successes in a sequence of n independent experiments, the pdf is given as $\binom{n}{x}p^x(1-p)^{n-x}$

(e) exponential distribution - the probability distribution that describes a process in which events occur continuously and independently at a constant average rate, the pdf is given as $\frac{1}{\beta}e^{\frac{-x}{\beta}}$

5. (extra credit) What is the relationship between the Bernoulli and binomial distributions?

Solution: The relationship between the Bernoulli and binomial distributions is that the Bernoulli distribution is a special case of the binomial distribution. This is because a single trial is conducted. This would mean n is 1 for a binomial distribution. In other words, a binomial random variable is a random variable that represents the number of successes in n successive independent trials of a Bernoulli distribution.

6. Suppose that random variable $X \sim N(2, 3)$. What is its expected value?

Solution: The expected value of $X \sim N(2, 3)$ is 2.

7. An *euclidean projection* of a d-dimensional point $y \in \mathbb{R}^d$ to a set $\mathcal{Z}$ is given by the following optimization problem:

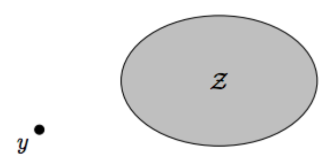
$$x^* = \arg\min_x ||x - y||_2^2, \text{ subject to: } x \in \mathcal{Z} \tag{2}$$

where $\mathcal{Z}$ is the feasible set, $||\cdot||_2$ is the ℓ_2-norm (euclidean) of a vector, and $x^* \in \mathbb{R}^d$ is the projected vector.

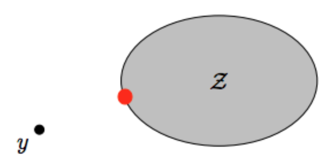
(a) What is x^* if $y = 1.1$ and $\mathcal{Z} = \mathbb{N}$, where $\mathbb{N}$ is the set of natural numbers?

Solution: We know that $y = 1.1$ and $\mathcal{Z} = \mathbb{N}$. To figure out what x^* is, I will plug in some natural numbers starting at 1 and increasing to see which values of x^* minimize the given optimization problem. First plugging 1 in, we see that the distance between these points is $\sqrt{(1-1.1)^2} = \sqrt{(-0.1)^2} = \sqrt{0.01} = 0.1$. Next plugging in 2, the distance is $\sqrt{(2-1.1)^2} = \sqrt{(0.9)^2} = \sqrt{0.81} = 0.9$. Continuing to see the pattern, plugging in 3, the distance is $\sqrt{(3-1.1)^2} = \sqrt{(1.9)^2} = \sqrt{3.61} = 1.9$. The natural numbers we are plugging in are continuously increasing and the results we are obtaining are also continuously increasing. Therefore, the first value we plugged in, 1, is the value of x^* that will minimize this optimization problem given. The smallest result we obtain is 0.1, therefore $x^* = 1$.

(b) Locate x^* in the following picture:



Solution: In this picture, x^* is located at the red dot.



8. Suppose that random variable Y has distribution:

$$p(Y = y) = \begin{cases} e^{-y} & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Verify that $\int_{y=-\infty}^{\infty} p(Y = y) = 1$.

Solution: To verify this integral we have:

$$\begin{aligned}
\int_{y=-\infty}^{\infty} p(Y=y) &= \int_{-\infty}^{0} p(Y=y) + \int_{0}^{\infty} p(Y=y) \\
&= \int_{-\infty}^{0} 0 + \int_{0}^{\infty} e^{-y} \\
&= 0 + \int_{0}^{\infty} e^{-y} \\
&= -e^{-y}\Big|_0^\infty \\
&= -e^{-\infty} - (-e^{-(0)}) \\
&= 0 + 1 \\
&= 1.
\end{aligned}$$

Thus we have shown that $\int_{y=-\infty}^{\infty} p(Y=y) = 1$.

(b) What is $\mu_Y = E[Y] = \int_{y=-\infty}^{\infty} p(Y=y)ydy$; that is, the expected value of Y?

Solution: To find μ_Y we have:

$$\begin{aligned}
\mu_Y &= E[Y] \\
&= \int_{y=-\infty}^{\infty} p(Y=y)ydy \\
&= \int_{-\infty}^{0} p(Y=y)ydy + \int_{0}^{\infty} p(Y=y)ydy \\
&= \int_{-\infty}^{0} (0)(y)dy + \int_{0}^{\infty} (e^{-y})(y)dy \\
&= \int_{-\infty}^{0} 0dy + \int_{0}^{\infty} ye^{-y}dy \\
&= 0 + \int_{0}^{\infty} ye^{-y}dy \\
&= \int_{0}^{\infty} ye^{-y}dy.
\end{aligned}$$

Using integration by parts and our previous solution, we can solve this integral as follows:

$$\begin{aligned}
\mu_Y &= \int_{0}^{\infty} ye^{-y}dy \\
&= -ye^{-y}\Big|_0^\infty + \int_{0}^{\infty} e^{-y}dy \\
&= -ye^{-y} - e^{-y}\Big|_0^\infty \\
&= e^{-y}(-y-1)\Big|_0^\infty \\
&= [e^{-\infty}(-\infty-1)] - [e^{-(0)}(-(0)-1)] \\
&= 0 - [1(-1)] \\
&= 1.
\end{aligned}$$

Thus we have $\mu_Y = 1$.

(c) What is $\sigma^2 = \text{Var}[Y] = \int_{y=-\infty}^{\infty} p(Y=y)(y-\mu_Y)^2dy$; that is, the variance of Y?

Solution: To find σ^2 we have:

$$\begin{aligned}
\sigma^2 &= \mathrm{Var}[Y] \\
&= \int_{y=-\infty}^{\infty} p(Y=y)(y-\mu_Y)^2 dy \\
&= \int_{-\infty}^{0} p(Y=y)(y-\mu_Y)^2 dy + \int_{0}^{\infty} p(Y=y)(y-\mu_Y)^2 dy \\
&= \int_{-\infty}^{0} (0)(y-\mu_Y)^2 dy + \int_{0}^{\infty} (e^{-y})(y-\mu_Y)^2 dy \\
&= 0 + \int_{0}^{\infty} (e^{-y})(y-\mu_Y)^2 dy \\
&= \int_{0}^{\infty} (e^{-y})(y-\mu_Y)^2 dy.
\end{aligned}$$

In the previous part we found that $\mu_Y = 1$, so now we can plug that into our integral as follows:

$$\begin{aligned}
\sigma^2 &= \int_{0}^{\infty} (e^{-y})(y-\mu_Y)^2 dy \\
&= \int_{0}^{\infty} (e^{-y})(y-1)^2 dy \\
&= \int_{0}^{\infty} (e^{-y})(y^2-2y+1) dy \\
&= \int_{0}^{\infty} y^2e^{-y} - 2ye^{-y} + e^{-y} dy \\
&= \int_{0}^{\infty} y^2e^{-y} dy + \int_{0}^{\infty} -2ye^{-y} dy + \int_{0}^{\infty} e^{-y} dy \\
&= \int_{0}^{\infty} y^2e^{-y} dy - 2\int_{0}^{\infty} ye^{-y} dy + \int_{0}^{\infty} e^{-y} dy.
\end{aligned}$$

Using parts a and b, we can plug our results into the second two integrals we have. For the first integral, we can solve it using integration by parts as follows:

$$\begin{aligned}
\sigma^2 &= \int_{0}^{\infty} y^2e^{-y} dy - 2\int_{0}^{\infty} ye^{-y} dy + \int_{0}^{\infty} e^{-y} dy \\
&= \left(-y^2e^{-y}\Big|_0^\infty + 2\int_{0}^{\infty} ye^{-y} dy \right) - 2\left(e^{-y}(-y-1)\Big|_0^\infty \right) + \left(-e^{-y}\Big|_0^\infty \right) \\
&= \left((0-0) + 2(e^{-y}(-y-1)\Big|_0^\infty \right) - 2(1) + 1 \\
&= 2(1) - 2(1) + 1 \\
&= 2 - 2 + 1 \\
&= 1.
\end{aligned}$$

Therefore we have $\sigma^2 = 1$.

(d) What is $E[Y|Y \geq 10]$; that is, the expected value of Y, given that (or conditioned on) $Y \geq 10$?

Solution: To find $E[Y|Y \geq 10]$ we have:

$$\begin{aligned}
E[Y|Y \geq 10] &= \int_{10}^{\infty} p(Y=y) y dy \\
&= \int_{10}^{\infty} (e^{-y})(y) dy \\
&= \int_{10}^{\infty} ye^{-y} dy.
\end{aligned}$$

From here we can use our results from above to find:

$$
\begin{aligned}
E[Y|Y \geq 10] &= \int_{10}^{\infty} ye^{-y} dy \\
&= e^{-y}(-y-1)\Big|_{10}^{\infty} \\
&= [e^{-\infty}(-\infty - 1)] - [e^{-10}(-10-1)] \\
&= 0 - [e^{-10}(-11)] \\
&= 11e^{-10}.
\end{aligned}
$$

Therefore, we have $E[Y|Y \geq 10] = 11e^{-10}$.