# Jeet Shah

G41262537

Project – 3 Report

Part-1:

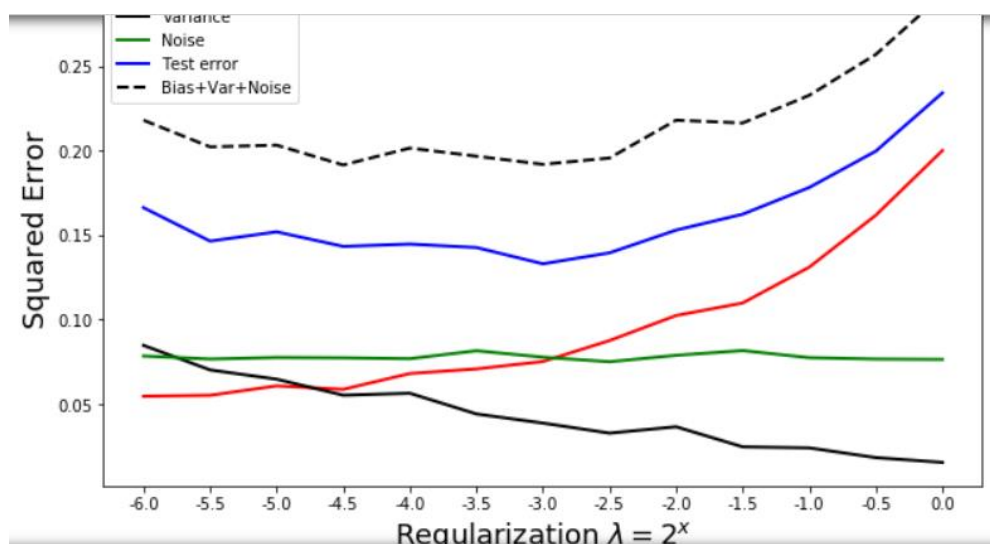**Link: https://github.com/jeetsj/ML_GWU/blob/master/python_Bonus_BV.ipynb**

Here, we are asked to decompose error into bias, variance and noise when regularization constant increases.

Here, first there is computeybar() function that is modified. It returns vector of the expected labels for the given input(array of n vectors).

In the below diagram misclassified points are highlighted.
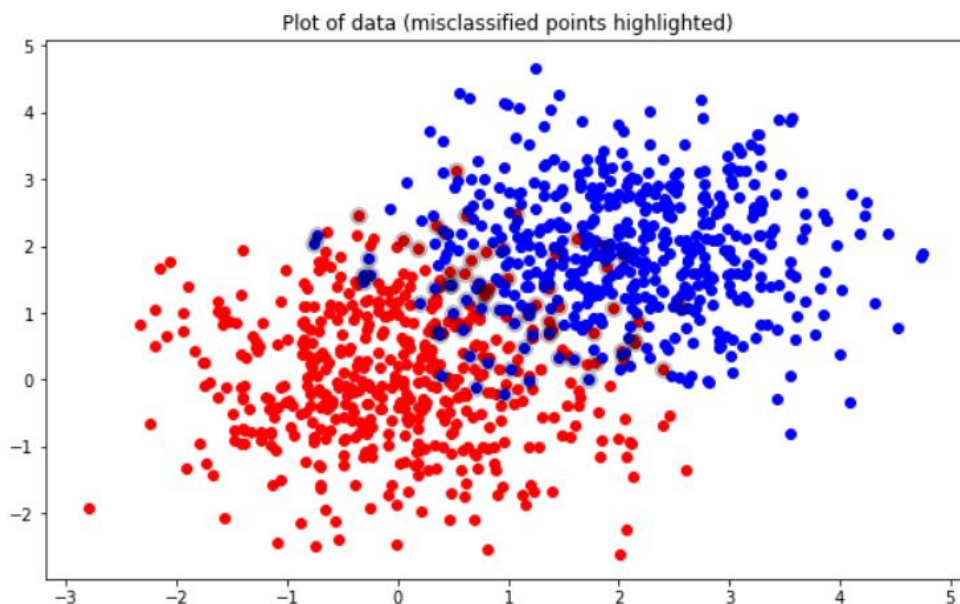
After modification, this is how bias, variance and noise will change with regularization.
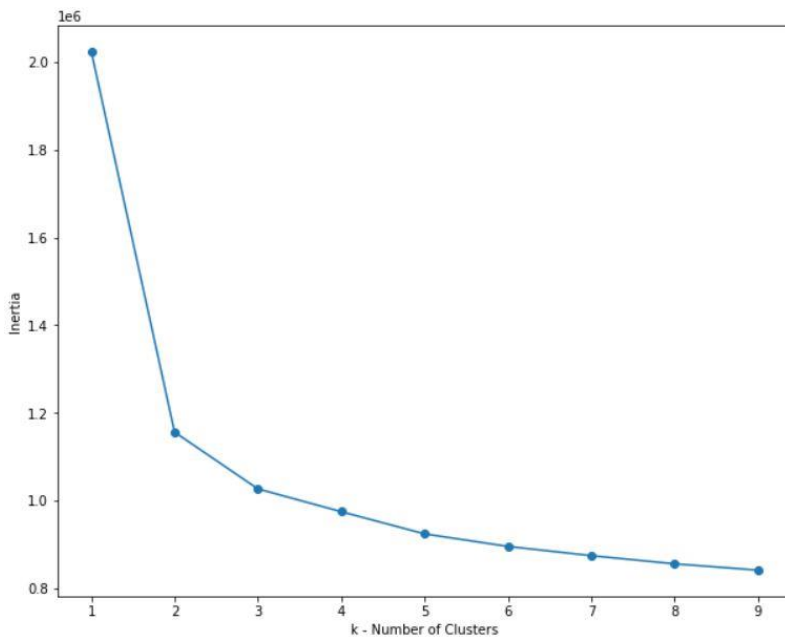
Part – 2:

Question – 1:

Link: https://github.com/jeetsj/ML_GWU/blob/master/Human_Activity.ipynb

For first question, we are given Human Activity Recognition dataset. Here, there are 6 labels: WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING.
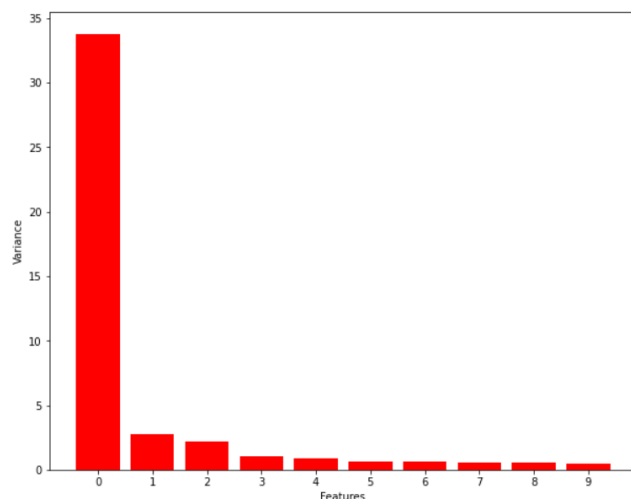
As shown in below diagram, k (number of clusters) = 2 is ideal.

This is the result for the k means clustering(k=2).

```
Homogeneity_score: 0.977400570620251
Completeness_score: 0.977696977792435
v_measure_score: 0.977548751737589
Inertia: 1156484.008987799
silhouette_score: 0.38990168994287766
```

PCA:

Here, 1 feature seems suitable.

We can see that Inertia and silhouette score has increased a lot with PCA.

```
Homogeneity_score: 0.97567693774092
Completeness_score: 0.9761994671205917
v_measure_score: 0.9759381324885832
Inertia: 168716.31233525547
silhouette_score: 0.7941361821630101
```
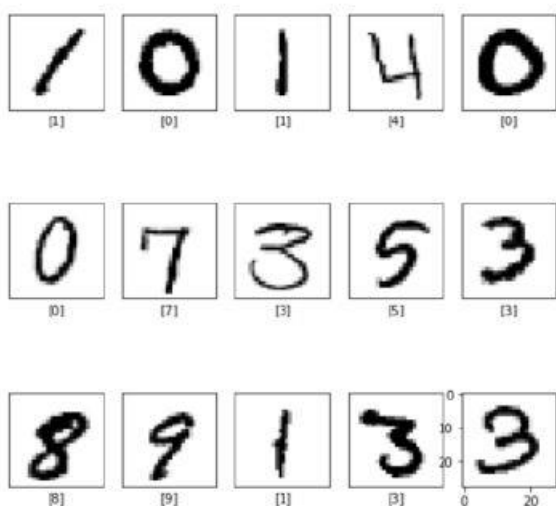
**Question-2:**

Here, we are given MNIST dataset.

Dataset Details:
In the training set, there are 785 colums. First column represents label and rest every column has a name pixelx, where x is an integer between 0 and 783, inclusive. We are given image data which is of 28*28 (784) pixels. Pixels value are from 0 to 255.
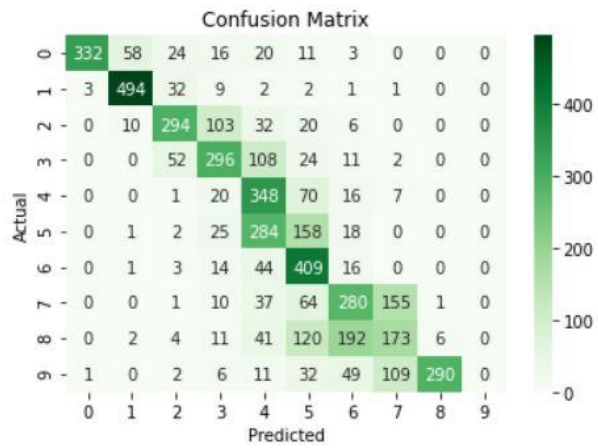
One example of the numbers is shown below.

Using KNN (I have taken 4 neighbors):
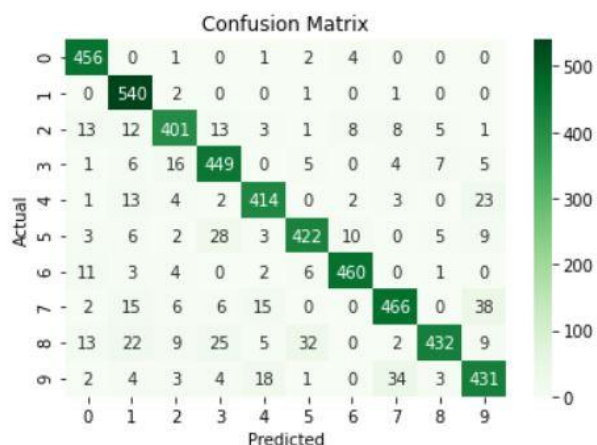
Using Random Forest:

Confusion Matrix

KNN seems to be more efficient as it is more accurate. The classification report is also shown in "Notebook".