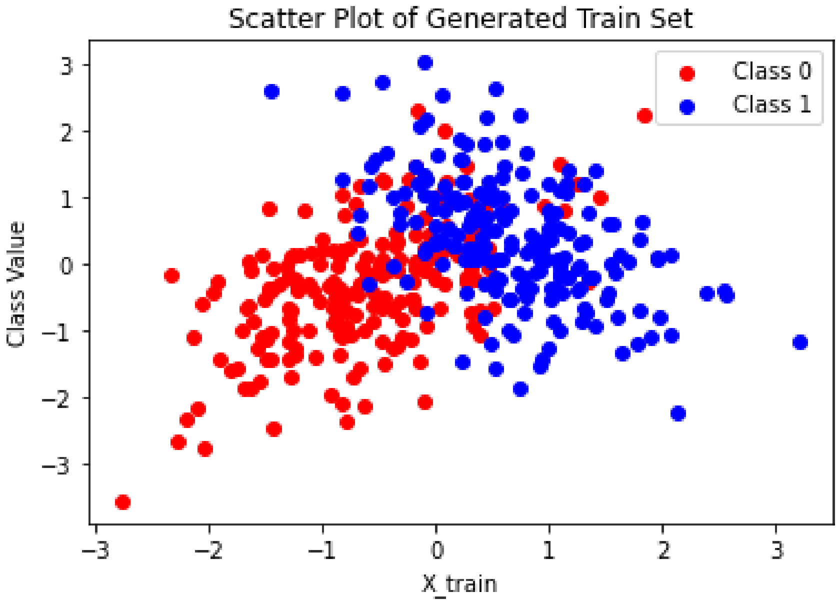
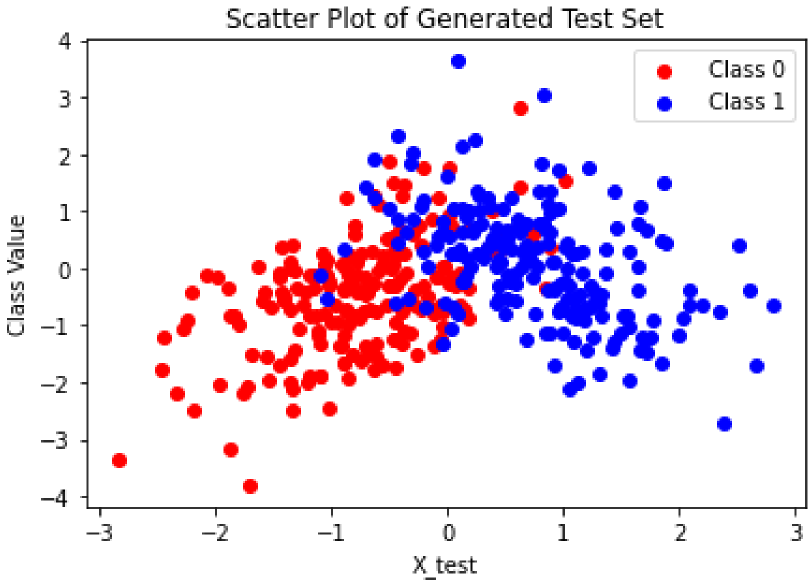
Project 1

UIN: 01183457

Part 01

Task :01 Plotting Scatter Plot on Generated Train and Test Data





Task 2: Bayes Classifier Performance

Added small value on the diagonal of covariance matrix= 0.006

Data	Training Accuracy (%)	Testing Accuracy (%)
Generated Data	86.0 %	87.5%
Zip Code Data	91.3%	88.4 %

Task 3: Naïve Bayes Classifier Performance

Data	Training Accuracy (%)	Testing Accuracy (%)
Generated Data	85.0 %	88.0 %
Zip Code Data	89.13 %	86.0 %

Task 4: Non-parametric Estimation of Gaussian Kernel

The value of h, is =0.5

Data	Training Accuracy (%)	Testing Accuracy (%)
Generated Data	86.0 %	88.25 %
Zip Code Data	89.1 %	87.23 %

The value of h, is =0.7

Data	Training Accuracy (%)	Testing Accuracy (%)
Generated Data	85.75 %	88.0 %
Zip Code Data	87.13 %	84.83 %

The value of h, is =0.9

Data	Training Accuracy (%)	Testing Accuracy (%)
Generated Data	85.5 %	87.5 %
Zip Code Data	84. 3 %	81.89 %

Task 5: K-Nearest Neighbor

Generated Data

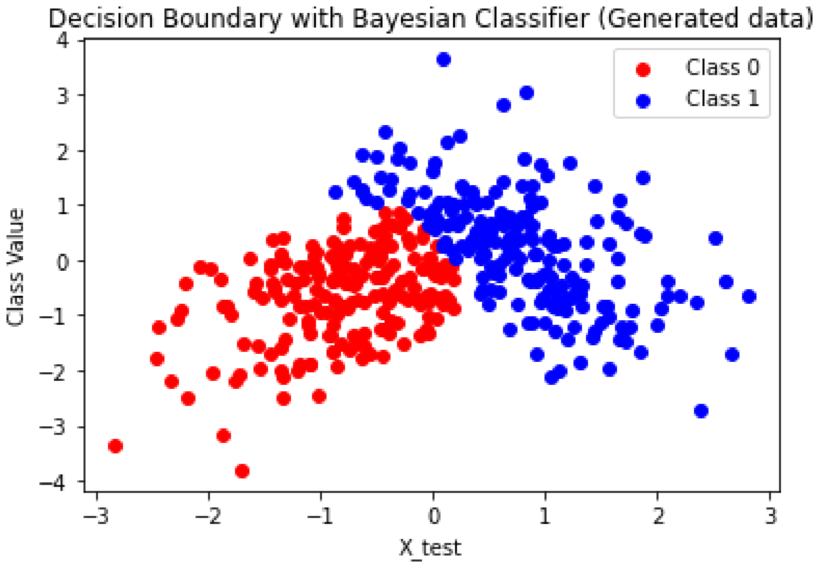
K-values	Testing Accuracy
3	87.75 %
5	87.75%
7	87.50 %

Zip-Code Data

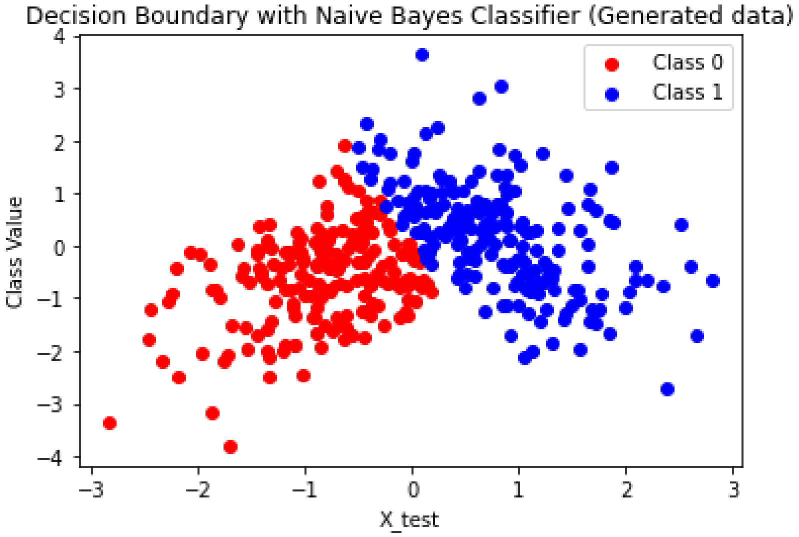
K-values	Testing Accuracy
3	99.33 %
5	99.0 %
7	99.0 %

Task 6: Plotting Decision Boundary on Generated Data

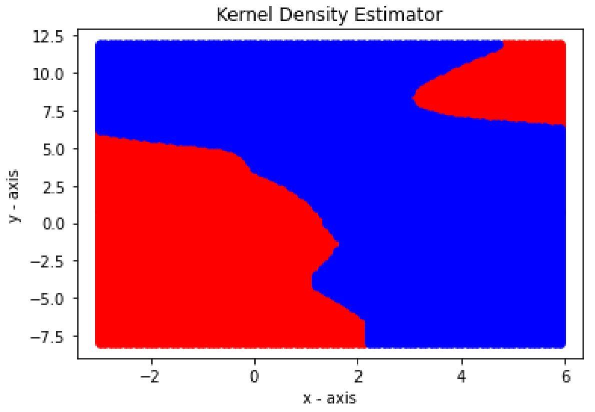
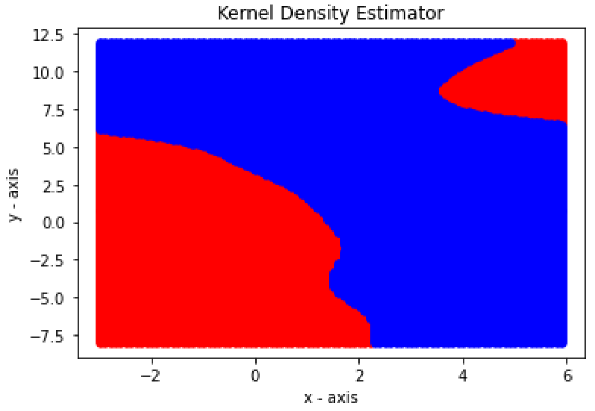
Bayes Classifier:

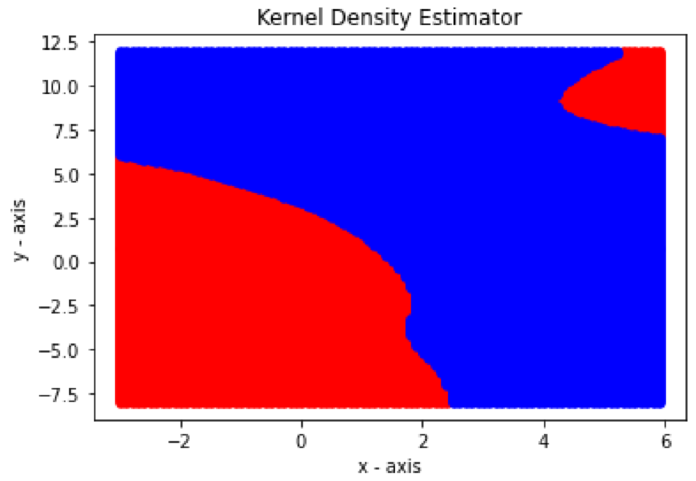


Naïve Bayes Classifier:

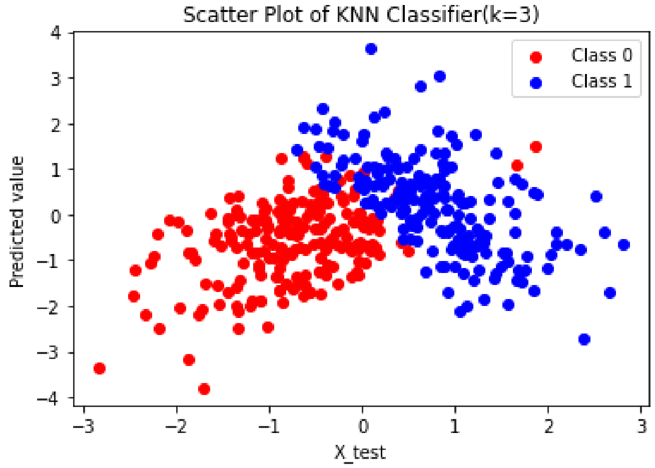
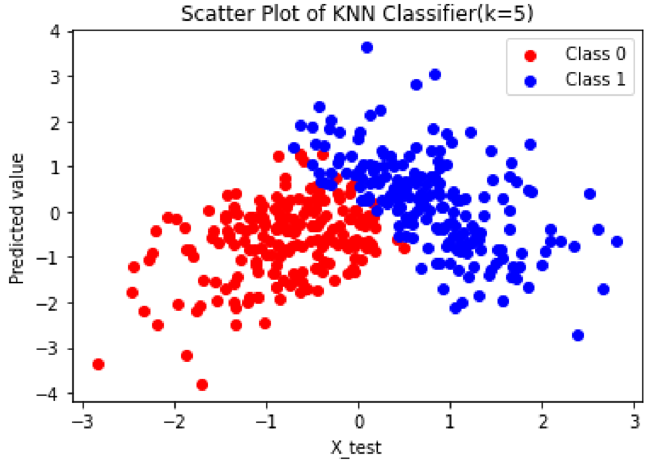

Decision Boundary with Naive Bayes Classifier (Generated data)

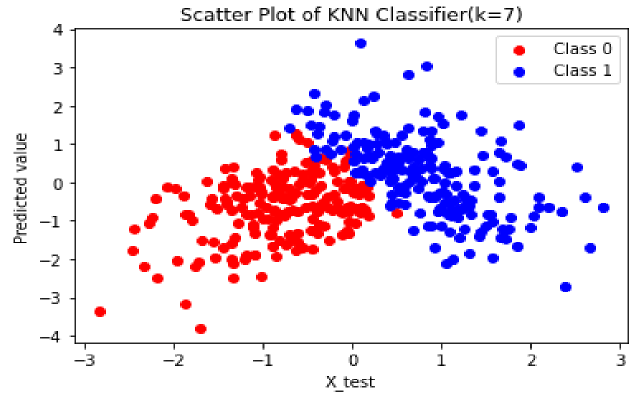
Non-Parametric Estimation:

H=0.5	 Kernel Density Estimator
H=0.7	 Kernel Density Estimator

H=0.9	

K-Nearest Neighbor:

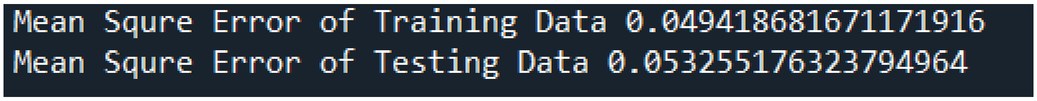
K=3	
K=5	

K=7	

Part 02

Task :01

Training Mean Square Error (MSE)	0.049
Testing Mean Square Error (MSE)	0.053



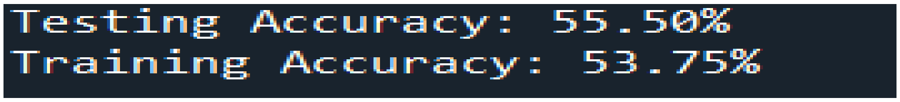
```
Mean Squre Error of Training Data 0.049418681671171916
Mean Squre Error of Testing Data 0.053255176323794964
```

Task: 02

Generated Data:

Using bias term (Adding a column of 1’s to Train and Test Data)

Training Accuracy (%)	53.75 %
Testing Accuracy (%)	55.50 %



```
Testing Accuracy: 55.50%
Training Accuracy: 53.75%
```

Comparison:

Classifiers	**Testing Accuracy (%)**
Bayes Classifier	87.5 %
Naïve Bayes Classifier	88.0 %
Non-Parametric Bayes Classifier	87.5 % (h=0.9)
K-Nearest Neighbor	87.75% (k=3)
Linear Classifier	55.50 %

Task: 03

Zip-code Data:

Regularization Co-efficient	Testing Accuracy (%)
0.01	41.53 %
0.1	41.50 %
0.5	41.50 %
1.0	41.50 %
5.0	41.50 %
10.0	41.50 %

There is no difference in testing accuracy after adding different values of regularization co-efficient to the diagonal of co-variance matrix in case of zip-code data in case of linear classification which is based on calculating weights from train data. Although, there is little difference in testing accuracy between regularization co-efficient of 0.01 and other co-efficient values.

Comparison:

Classifiers	**Testing Accuracy (%)**
Bayes Classifier	88.4 %
Naïve Bayes Classifier	86.0 %
Non-Parametric Bayes Classifier	81.89 % (h=0.9)
K-Nearest Neighbor	99.33% (k=3)
Linear Classifier	41.53 % (regularization co-efficient=0.01)

It is observed from the acquired testing accuracy on zip-code and generated data, the linear classifier does not perform well compared to the other classification model. In case of generated data, Naïve Bayes classifier perform the best with testing accuracy of 88.0 % and in case of zip-code data, K-Nearest Classifier perform the best with classification accuracy of 99.33%.

References:

1. Website Link: https://www.codegrepper.com/code-examples/python/how+to+plot+two+different+class+in+different+colour+in+python

2. Website Link: https://sophiamyang.github.io/DS/optimization/multiclass-logistic/multiclass-logistic.html

3. Website Link: https://github.com/knowthebird/KernelDensityEstimator/blob/master/kerneldesnityestimator.py

4. Website Link: https://github.com/abhilampard/Multiple-Linear-Regression-From-Scratch/blob/master/Multiple_Linear_Regression.ipynb

5. Website Link: https://github.com/abhilampard/Multiple-Linear-Regression-From-Scratch/blob/master/Multiple_Linear_Regression.ipynb