Kolby Sarson
104232327

Assignment 1

Running SciKit-Learn

To run the classifiers on SciKit-Learn I first imported 'LinearDiscriminantAnalysis' and 'QuadraticDiscriminantAnalysis' from 'sklearn.discriminant_analysis' as well as 'GaussianNB' from 'sklearn.naive_bayes' and 'svm' from 'sklearn.' I then created an array for the four classifiers in order to loop through each classifier for each dataset. Next, I split the dataset into random train and test subsets. After that I standardized the dataset so that it would perform correctly. Finally, I had an if statement for each classifier and ran each of the four in a loop, one by one.

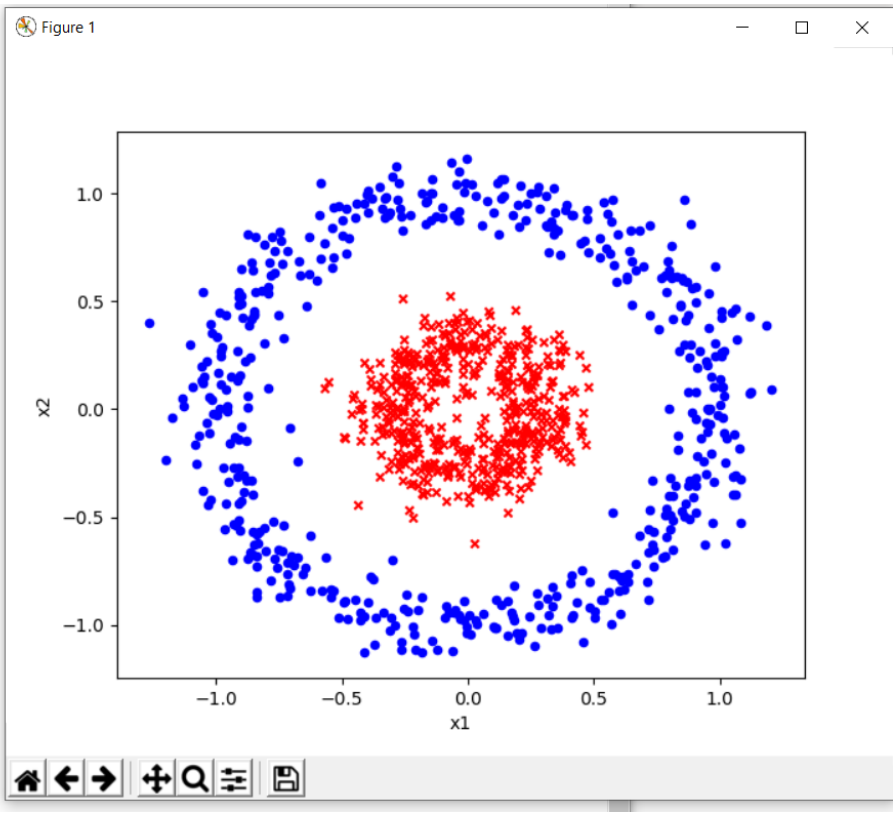
Kolby Sarson
104232327

```
LDA for circles0.3.csv
Confusion Matrix
[[52 45]
 [90 13]]
Accuracy 0.325
Positive Predictive Value: 0.54
Negative Predictive Value: 0.13
Sensitivity: 0.37
Specificity: 0.22

QDA for circles0.3.csv
Confusion Matrix
[[ 97   0]
 [  1 102]]
Accuracy 0.995
Positive Predictive Value: 1.00
Negative Predictive Value: 0.99
Sensitivity: 0.99
Specificity: 1.00

Bayes for circles0.3.csv
Confusion Matrix
[[ 97   0]
 [  1 102]]
Accuracy 0.995
Positive Predictive Value: 1.00
Negative Predictive Value: 0.99
Sensitivity: 0.99
Specificity: 1.00

SVM for circles0.3.csv
Confusion Matrix
[[ 97   0]
 [  0 103]]
Accuracy 1.0
Positive Predictive Value: 1.00
Negative Predictive Value: 1.00
Sensitivity: 1.00
Specificity: 1.00
```



circles0.3

- LDA
  - The accuracy is poor because no matter how you divide the graph linearly, the classification will not be able to separate classes effectively
- QDA
  - The accuracy is strong because QDA will produce an ellipse that would classify most data efficiently
- Bayes
  - The accuracy is strong because Gaussian Naïve Bayes will be able to accurately classify the data based on probability since the datasets are very separated
- SVM
  - The accuracy is perfect because the 2D data, without overlapping data between the two classes, can be projected into 3D creating a perfectly separating hyperplane

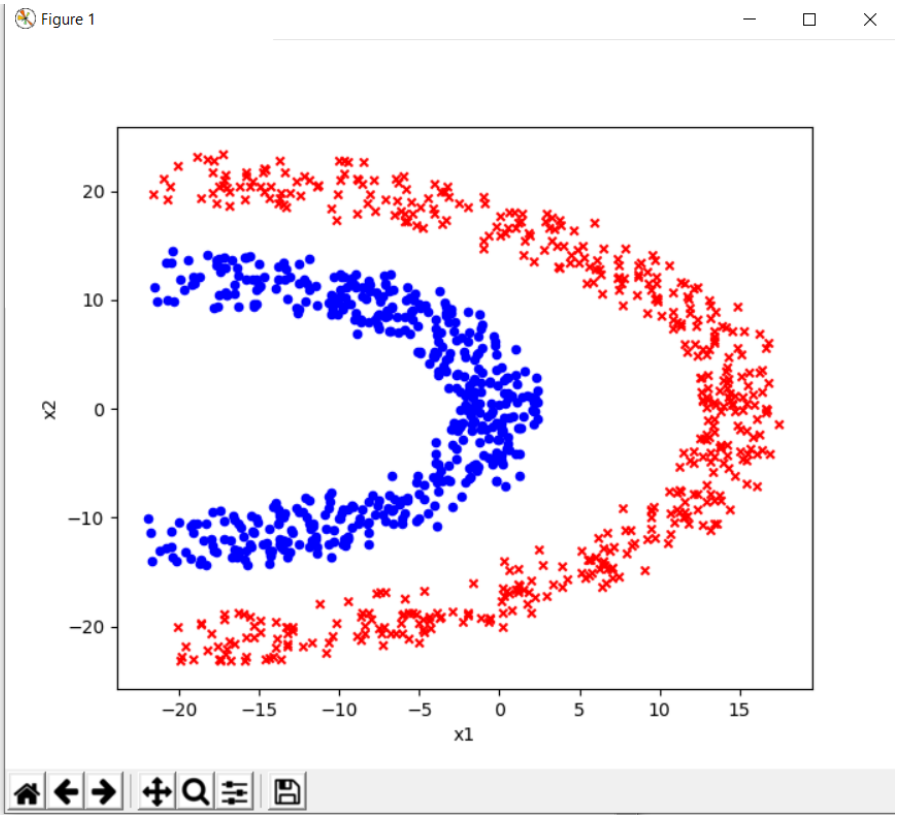
Kolby Sarson
104232327

```
LDA for halfkernel.csv
Confusion Matrix
[[71 27]
 [34 68]]
Accuracy 0.695
Positive Predictive Value: 0.72
Negative Predictive Value: 0.67
Sensitivity: 0.68
Specificity: 0.72

QDA for halfkernel.csv
Confusion Matrix
[[97  1]
 [10 92]]
Accuracy 0.945
Positive Predictive Value: 0.99
Negative Predictive Value: 0.90
Sensitivity: 0.91
Specificity: 0.99

Bayes for halfkernel.csv
Confusion Matrix
[[98  0]
 [11 91]]
Accuracy 0.945
Positive Predictive Value: 1.00
Negative Predictive Value: 0.89
Sensitivity: 0.90
Specificity: 1.00

SVM for halfkernel.csv
Confusion Matrix
[[ 98   0]
 [  0 102]]
Accuracy 1.0
Positive Predictive Value: 1.00
Negative Predictive Value: 1.00
Sensitivity: 1.00
Specificity: 1.00
```



halfkernel

- LDA
  - The accuracy is alright since a linear separation should be able to separate most of class 1 from approximately half of class 0
- QDA
  - The accuracy is strong since the shapes of the classes are close to parabolic, thus a parabolic line would be able to separate the classes effectively
- Bayes
  - The accuracy is strong because Gaussian Naïve Bayes will be able to accurately classify the data based on probability since the datasets are very separated
- SVM
  - The accuracy is perfect because the 2D data, without overlapping data between the two classes, can be projected into 3D creating a perfectly separating hyperplane

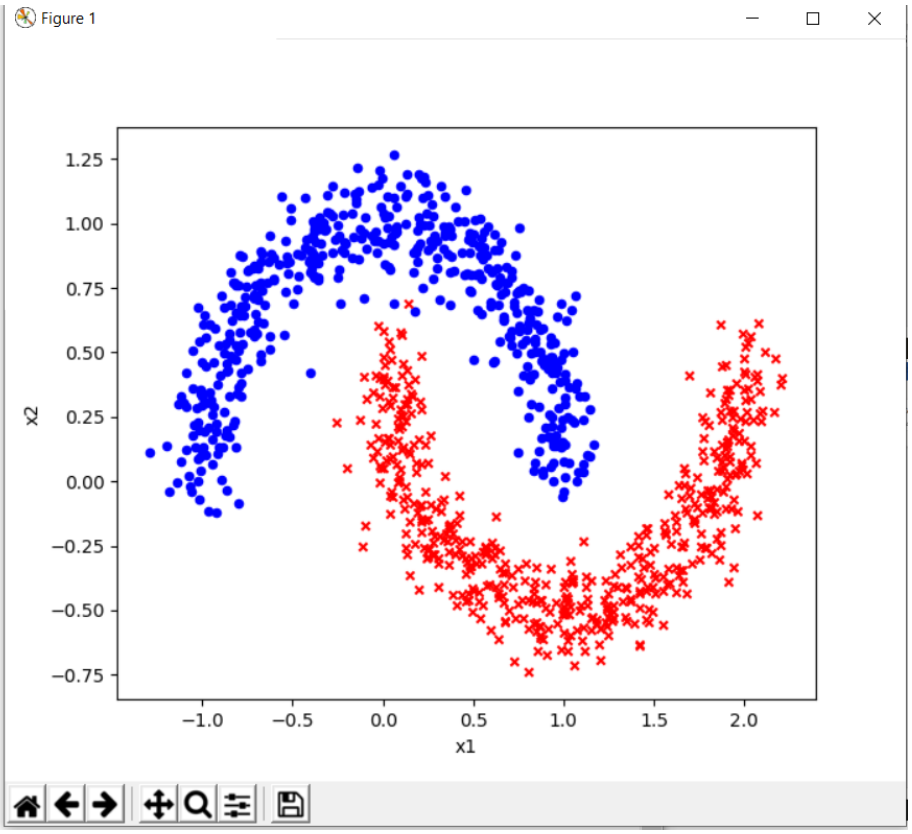
Kolby Sarson
104232327

```
LDA for moons1.csv
Confusion Matrix
[[89 11]
 [12 88]]
Accuracy 0.885
Positive Predictive Value: 0.89
Negative Predictive Value: 0.88
Sensitivity: 0.88
Specificity: 0.89

QDA for moons1.csv
Confusion Matrix
[[89 11]
 [12 88]]
Accuracy 0.885
Positive Predictive Value: 0.89
Negative Predictive Value: 0.88
Sensitivity: 0.88
Specificity: 0.89

Bayes for moons1.csv
Confusion Matrix
[[89 11]
 [12 88]]
Accuracy 0.885
Positive Predictive Value: 0.89
Negative Predictive Value: 0.88
Sensitivity: 0.88
Specificity: 0.89

SVM for moons1.csv
Confusion Matrix
[[100   0]
 [  0 100]]
Accuracy 1.0
Positive Predictive Value: 1.00
Negative Predictive Value: 1.00
Sensitivity: 1.00
Specificity: 1.00
```



moons1

- LDA
  - The accuracy is strong because a linear separation would be able to differentiate most of class 0 from most of class 1
- QDA
  - The accuracy is strong (the same as LDA) because QDA would also produce the same line as in LDA
- Bayes
  - The accuracy is strong since the datasets are very separated, with only one area where the two class plots come quite close which slightly reduces probabilities and the accuracy
- SVM
  - The accuracy is perfect because the 2D data, without overlapping data between the two classes, can be projected into 3D creating a perfectly separating hyperplane

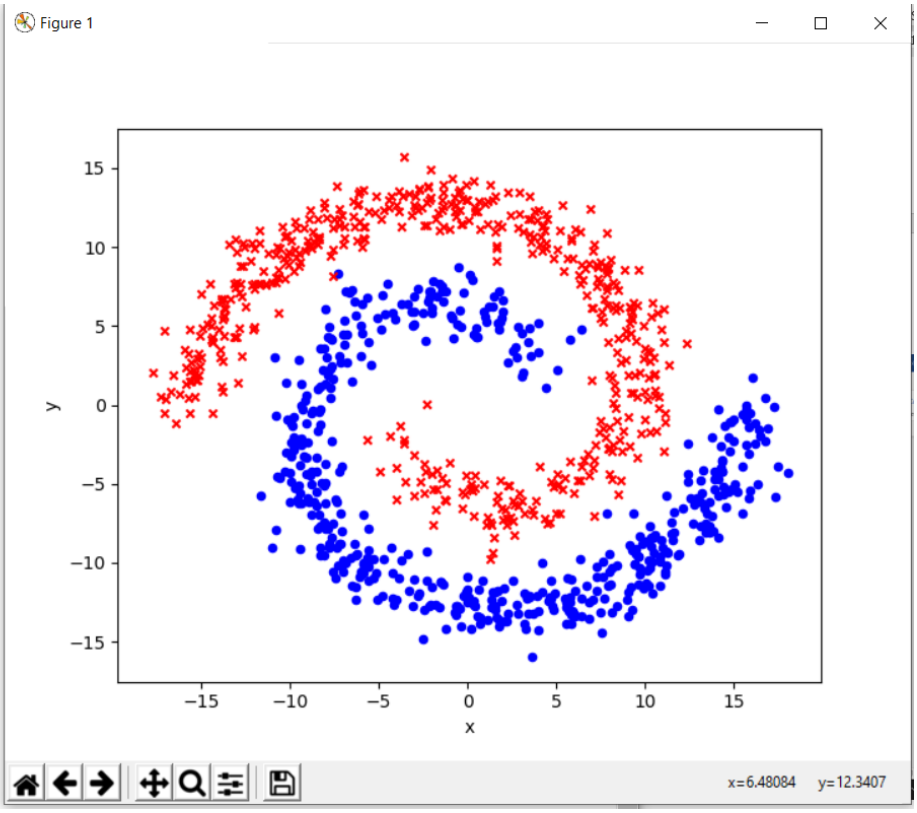
Kolby Sarson
104232327

```
LDA for spiral1.csv
Confusion Matrix
[[68 26]
 [32 74]]
Accuracy 0.71
Positive Predictive Value: 0.72
Negative Predictive Value: 0.70
Sensitivity: 0.68
Specificity: 0.74

QDA for spiral1.csv
Confusion Matrix
[[68 26]
 [32 74]]
Accuracy 0.71
Positive Predictive Value: 0.72
Negative Predictive Value: 0.70
Sensitivity: 0.68
Specificity: 0.74

Bayes for spiral1.csv
Confusion Matrix
[[64 30]
 [32 74]]
Accuracy 0.69
Positive Predictive Value: 0.68
Negative Predictive Value: 0.70
Sensitivity: 0.67
Specificity: 0.71

SVM for spiral1.csv
Confusion Matrix
[[ 94   0]
 [  2 104]]
Accuracy 0.99
Positive Predictive Value: 1.00
Negative Predictive Value: 0.98
Sensitivity: 0.98
Specificity: 1.00
```



spiral1

- LDA
  - The accuracy is alright since a linear separation should be able to separate a large portion (but not the majority) of each class from one another
- QDA
  - The accuracy is alright (the same as LDA) because QDA would also produce the same line as in LDA
- Bayes
  - The accuracy is alright because the class plots are sufficiently separated, but there are many areas where the two classes come quite close to each other, affecting the probabilities and the accuracy
- SVM
  - The accuracy is near perfect because the 2D data, with very minimal overlapping data between the two classes, can be projected into 3D creating a near-perfectly separating hyperplane

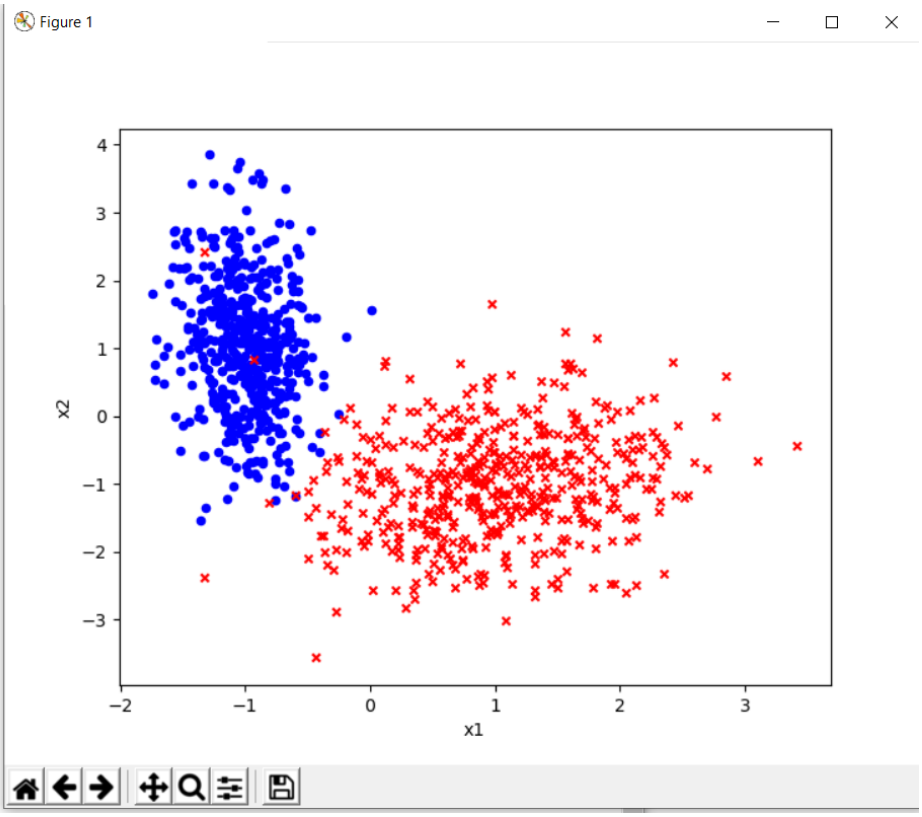
Kolby Sarson
104232327

```
LDA for twogaussians33.csv
Confusion Matrix
[[ 97   0]
 [  3 100]]
Accuracy 0.985
Positive Predictive Value: 1.00
Negative Predictive Value: 0.97
Sensitivity: 0.97
Specificity: 1.00

QDA for twogaussians33.csv
Confusion Matrix
[[ 97   0]
 [  1 102]]
Accuracy 0.995
Positive Predictive Value: 1.00
Negative Predictive Value: 0.99
Sensitivity: 0.99
Specificity: 1.00

Bayes for twogaussians33.csv
Confusion Matrix
[[ 96   1]
 [  2 101]]
Accuracy 0.985
Positive Predictive Value: 0.99
Negative Predictive Value: 0.98
Sensitivity: 0.98
Specificity: 0.99

SVM for twogaussians33.csv
Confusion Matrix
[[ 96   1]
 [  2 101]]
Accuracy 0.985
Positive Predictive Value: 0.99
Negative Predictive Value: 0.98
Sensitivity: 0.98
Specificity: 0.99
```



twogaussians33

- LDA
  - The accuracy is near perfect since a linear separation would be able to differentiate the majority of class 0 from the majority of class 1 (with minimal points overlapping)
- QDA
  - The accuracy is near perfect since a parabolic shape or ellipse could entirely (or near entirely) encompass one class with minimal overlapping points from the other class
- Bayes
  - The accuracy is near perfect since the datasets are very separated, with only a few points overlapping between classes, and only one area where the two class plots come quite close which slightly reduces probabilities and the accuracy
- SVM
  - The accuracy is near perfect because the 2D data, with very minimal overlapping data between the two classes, can be projected into 3D creating a near-perfectly separating hyperplane

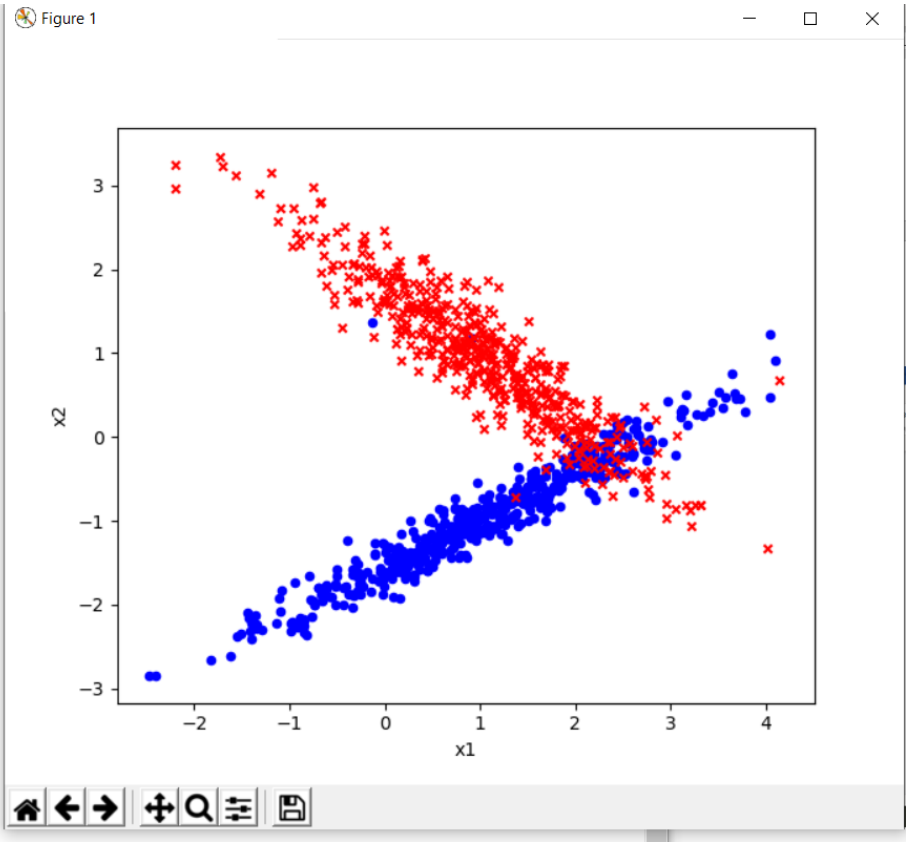
Kolby Sarson
104232327

```
LDA for twogaussians42.csv
Confusion Matrix
[[93 10]
 [12 85]]
Accuracy 0.89
Positive Predictive Value: 0.90
Negative Predictive Value: 0.88
Sensitivity: 0.89
Specificity: 0.89

QDA for twogaussians42.csv
Confusion Matrix
[[102   1]
 [  7  90]]
Accuracy 0.96
Positive Predictive Value: 0.99
Negative Predictive Value: 0.93
Sensitivity: 0.94
Specificity: 0.99

Bayes for twogaussians42.csv
Confusion Matrix
[[93 10]
 [11 86]]
Accuracy 0.895
Positive Predictive Value: 0.90
Negative Predictive Value: 0.89
Sensitivity: 0.89
Specificity: 0.90

SVM for twogaussians42.csv
Confusion Matrix
[[103   0]
 [ 10  87]]
Accuracy 0.95
Positive Predictive Value: 1.00
Negative Predictive Value: 0.90
Sensitivity: 0.91
Specificity: 1.00
```



twogaussians42

- LDA
  - The accuracy is strong because a linear separation would be able to differentiate most of class 0 from most of class 1
- QDA
  - The accuracy is strong because a parabolic shape or an ellipse could entirely (or near entirely) encompass one class with only a small portion overlapping points from the other class
- Bayes
  - The accuracy is strong since the datasets are very separated, with only one area where a small portion of points are overlapping between classes, which reduces probabilities and the accuracy
  -
- SVM
  - The accuracy is strong because the 2D data, with only one area where a small portion of points are overlapping between the two classes, can be projected into 3D creating a strong separating hyperplane