Advanced Topics in Artificial Intelligence

Assignment - 1

Submitted To
Dr. Luis Rueda

"As students of the University of Windsor, we pledge to pursue all endeavours with honour and integrity, and will not tolerate or engage in academic or personal dishonesty. We confirm that we have not received any unauthorized assistance in preparing for or writing this assignment. We acknowledge that a mark of 0 may be assigned for the copied work."

– **Jaskaran Singh Luthra 110090236**

Team Members and Student IDs

S No.	Team Member Name	Student ID
1.	Jaskaran Singh Luthra	110090236
2.	Harjot Singh Saggu	110093636
3.	Bikramjeet Singh	110099784

We used a 5 step process to find the measures on the classifiers for the given datasets:

1. Import the necessary libraries like “scikitlearn”
2. Load and prepare the dataset.
3. Initialize and fit the classifier model.
4. Make predictions on the dataset.
5. Calculate the measures of the model.

The measure for the classifiers shows the following readings:

1. **For the Linear Discriminant Analysis:**
   Circles Dataset:
   Accuracy Score: 0.5
   PPV: 0.4918918918918919
   NPV: 0.5130434782608696
   Specificity: 0.38562091503267976
   Sensitivity: 0.6190476190476191
   Moons Dataset:
   Accuracy Score: 0.8733333333333333
   PPV: 0.8859060402684564
   NPV: 0.8609271523178808
   Specificity: 0.8843537414965986
   Sensitivity: 0.8627450980392157
   Half Kernel Dataset:
   Accuracy Score: 0.68
   PPV: 0.6956521739130435
   NPV: 0.6666666666666666
   Specificity: 0.72
   Sensitivity: 0.64
   Two Gaussians Dataset:
   Accuracy Score: 0.8866666666666667
   PPV: 0.9407407407407408
   NPV: 0.8424242424242424
   Specificity: 0.9455782312925171
   Sensitivity: 0.8300653594771242
2. **For Quadratic Discriminant Analysis:**
   Circles Dataset:
   Accuracy Score: 0.9933333333333333
   PPV: 1.0
   NPV: 0.9870967741935484
   Specificity: 1.0
   Sensitivity: 0.9863945578231292
   Moons Dataset:
   Accuracy Score: 0.8733333333333333
   PPV: 0.8859060402684564

NPV: 0.8609271523178808
Specificity: 0.8843537414965986
Sensitivity: 0.8627450980392157

Half Kernel Dataset:

Accuracy Score: 0.9366666666666666
PPV: 0.9781021897810219
NPV: 0.901840490797546
Specificity: 0.98
Sensitivity: 0.8933333333333333

Two Gaussian Dataset:

Accuracy Score: 0.94
PPV: 0.972027972027972
NPV: 0.910828025477707
Specificity: 0.9727891156462585
Sensitivity: 0.9084967320261438

3. **For Naïve Baize Analysis:**

Circles Dataset:

Accuracy Score: 0.9933333333333333
PPV: 1.0
NPV: 0.9870967741935484
Specificity: 1.0
Sensitivity: 0.9863945578231292

Moons Dataset:

Accuracy Score: 0.8733333333333333
PPV: 0.8859060402684564
NPV: 0.8609271523178808
Specificity: 0.8843537414965986
Sensitivity: 0.8627450980392157

Half Kernel Dataset:

Accuracy Score: 0.9533333333333334
PPV: 0.9927536231884058
NPV: 0.9197530864197531
Specificity: 0.9933333333333333
Sensitivity: 0.9133333333333333

Two Gaussian Dataset:

Accuracy Score: 0.8833333333333333
PPV: 0.927536231884058
NPV: 0.845679012345679
Specificity: 0.9319727891156463
Sensitivity: 0.8366013071895425

4. **For SVM Analysis:**

Circles Dataset:

Accuracy Score: 0.62
PPV: 0.5673469387755102
NPV: 0.8545454545454545
Specificity: 0.30718954248366015
Sensitivity: 0.9455782312925171

Moons Dataset:
- Accuracy Score: 0.86
- PPV: 0.87248322147651
- NPV: 0.847682119205298
- Specificity: 0.8707482993197279
- Sensitivity: 0.8496732026143791

Half Kernel Dataset:
- Accuracy Score: 0.7466666666666667
- PPV: 0.8363636363636363
- NPV: 0.6947368421052632
- Specificity: 0.88
- Sensitivity: 0.6133333333333333

Two Gaussian Dataset:
- Accuracy Score: 0.9133333333333333
- PPV: 0.9922480620155039
- NPV: 0.8538011695906432
- Specificity: 0.9931972789115646
- Sensitivity: 0.8366013071895425

Accuracy of every classifier and dataset on why a particular classifier performed good/bad on a particular dataset:

For Linear Discriminant Analysis

The Circles dataset exhibits higher degree of non-linearity and non-linear separability, which may impede the capacity of Quadratic Discriminant Analysis (QDA) to classify the data effectively. Conversely, the Moons dataset exhibits a more pronounced boundary between classes, which may enhance the classification performance of QDA. The Half Kernel dataset exhibits a degree of class overlap, which may impede the classification performance of QDA. Conversely, the Two Gaussian dataset exhibits well-defined class separation, which may enhance the classification performance of QDA.

For Quadratic Discriminant Analysis

The variation in accuracy among the datasets can be attributed to the complexity and separability of the classes. The Circles dataset exhibits a high degree of linear separability, whereas the Moons dataset displays a more intricate and non-linearly separable class distribution. The Half Kernel and Two Gaussian datasets exhibit intermediate levels of class separability. As a result, it is probable that the Quadratic Discriminant Analysis (QDA) classifier will exhibit superior performance on datasets with simpler class distributions. Furthermore, the performance of QDA is also influenced by the sample size, feature dimensionality and the level of noise present in the data.

For Naive Bayes

The variation in accuracy among the datasets can be attributed to the extent of feature independence in each dataset. The Naive Bayes classifier assumes independence among features, thus it performs optimally on datasets where this assumption holds true, such as the Circles dataset. Conversely, datasets with correlated features, such as the Moons and Two Gaussian

dataset, may result in a decrease in performance for the Naive Bayes classifier. Additionally, the sample size, feature dimension, and noise level of the data can also have an impact on the performance of the Naive Bayes algorithm.

For SVM

The possible explanation for the low accuracy score of the SVM classifier on the Circles dataset is its non-linearity which may impede the model's ability to find an optimal decision boundary. This is further evidenced by the low Specificity (0.307) and high Sensitivity (0.945) values, which suggest a high rate of false positive classifications. Conversely, the Moons dataset, although also non-linearly separable, presents a less complex decision boundary, allowing the SVM model to perform better. However, the sample size, feature dimension, and noise level of the data also can affect the performance of SVM.