Jon Organ

CS677

Homework 11, Cardiography

**Question 1:**

Excel file read...

Classes assigned...

**Question 2:**

1) Naive bayesian class labels predicted...

2) Naive bayesian accuracy: 78.17%

3) Naive bayesian confusion matrix:

Predicted 0 1

Actual

0 38 196

1 36 793

**Question 3:**

1) Logistic regression class labels predicted...

2) Logistic regression accuracy: 80.15%

3) Logistic regression confusion matrix:

Predicted 0 1

Actual

0 29 205

1 6 823

**Question 4:**

1) Decision tree class labels predicted...

2) Logistic regression accuracy: 88.52%

3) Logistic regression confusion matrix:

Predicted 0 1

Actual

0 184 50

1 72 757

**Question 5:**

1) Random forest hyper-parameters run...

2) Saving random forest scatter plot...

A graph with blue dots

Description automatically generated

Best N value: 9

Best d value: 5

3) Random forest accuracy with highest performing hyper-parameters: 89.37%

4) Random forest confusion matrix:

Predicted 0 1

Actual

0 166 68

1 45 784

**Question 6:**

1) SVM class labels predicted...

2) SVM accuracy:

SVM Linear: 78.55%

SVM Degree 2: 80.24%

SVM Gaussian: 88.24%

3) SVM confusion matrices:

SVM Linear:

Predicted 0 1

Actual

0 8 226

1 2 827

SVM Degree 2:

Predicted 0 1

Actual

0 36 198

1 12 817

SVM Gaussian:

Predicted 0 1

Actual

0 136 98

1 27 802

**Question 7:**

Method TP FP TN FN accuracy TPR TNR

Naive Bayesian 38 36 793 196 0.78 0.16 0.96

Logistic 29 6 823 205 0.8 0.12 0.99

Decision Tree 184 72 757 50 0.89 0.79 0.91

Random Forest 166 45 784 68 0.89 0.71 0.95

Linear SVM 8 2 827 226 0.79 0.03 1.0

Degree 2 SVM 36 12 817 198 0.8 0.15 0.99

Gaussian SVM 136 27 802 98 0.88 0.58 0.97

**Question 8:**

One way to find the importance of features is to create various test groups with different features and create confusion matrices from each run. The better the score from the testing groups, the more important the tested features are. Weights can be assigned based off the resulting scores.