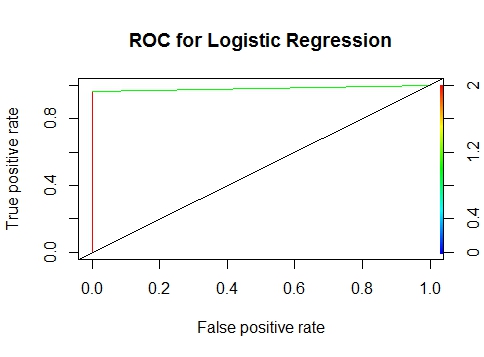
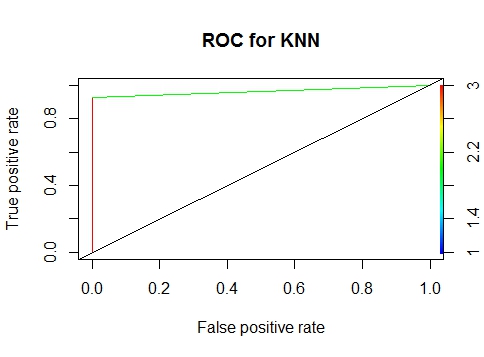
**REPORT- Performance comparison of Classifiers**

**Evaluation metric used : ROC curve plots**

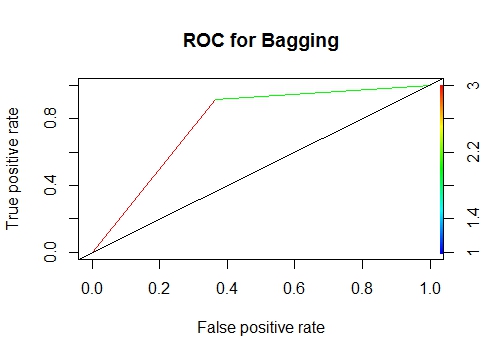
1. **LOGISTIC REGRESSION**

****

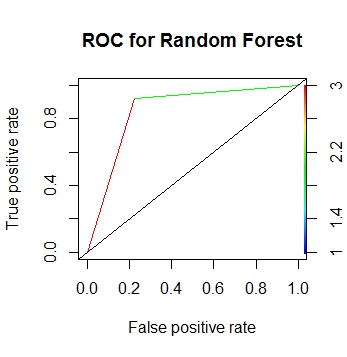
1. **K-NN**

****

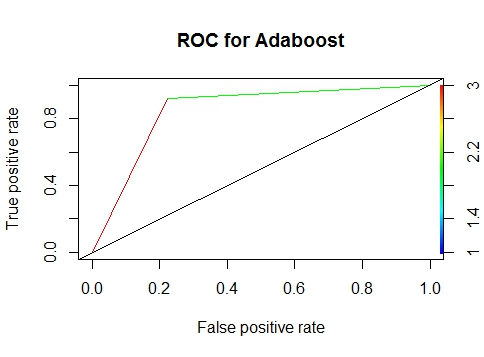
1. **BAGGING**

****

1. **RANDOM FOREST**

****

1. **BOOSTING**

****

**Results:**

Dataset: Ionosphere

Number of instances in dataset: 351

Number of attributes in dataset: 34

How many fold cross validation performed: 10

Packages used:

|  |  |  |
| --- | --- | --- |
| **Classifier** | **Package** | **Training function** |
| Logistic Regression | Stats | Glm with family = binomial |
| K-Nearest Neighbors | Class | Knn |
| Bagging | Adabag | Bagging |
| Random Forest | Randomforest | Randomforest |
| Boosting | Adabag | Boosting |

**Accuracy Variation:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Logistic Regression** | **K-NN** | **Bagging** | **Random Forest** | **Boosting** |
| 91.67 | 86.11 | 86.11 | 91.67 | 88.89 |
| 88.57 | 91.43 | 97.14 | 97.14 | 100.00 |
| 91.43 | 94.29 | 97.14 | 94.29 | 94.29 |
| 77.14 | 82.86 | 91.43 | 91.43 | 88.57 |
| 82.86 | 82.86 | 88.57 | 94.29 | 91.43 |
| 91.43 | 80.00 | 88.57 | 94.29 | 91.43 |
| 91.43 | 85.71 | 91.43 | 91.43 | 91.43 |
| 94.29 | 91.43 | 97.14 | 94.29 | 97.14 |
| 91.43 | 94.29 | 94.29 | 91.43 | 97.14 |
| 97.14 | 94.29 | 82.86 | 88.57 | 94.29 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Classifier** | **Technique – N fold** | **Average accuracy** | **Parameters** |
| Logistic Regression | 10 | 89.739 | Traindata, family=binomial |
| K-NN | 10 | 88.327 | K , probability |
| Bagging | 10 | 91.468 | Maximum depth |
| Random Forest | 10 | 92.883 | Ntree, mtree, importance, keep.forest=TRUE |
| Boosting | 10 | 93.461 | Iterations |

**Analysis:**

* 1. From the above results, Boosting is the best classifier with the accuracy of 93.461%.
  2. Accuracy varies between Random Forest, Boosting and Logistic Regression for every iteration.
  3. For k-NN classifier, we have considered k=9 neighbors.

Accuracy varies based on the no of neighbors.

Ex: When k=1 , accuracy will be around 94%

* 1. For Boosting, we have considered number of iterations = 25 , while we can spot the change in the ROC plot.
  2. If the maxdepth parameter of bagging changes, the accuracy will also change.
  3. The ROC plots give us the details of performance of every classifier.