|  |  |  |
| --- | --- | --- |
| **5-Fold Cross Validation** | | |
|  | Accuracy | Standard Deviation |
| Decision Tree | 79.3% | 1.6% |
| Bagging | 85.3% | 0.8% |
| Random Forest | 85.3% | 2.0% |
| Boosting | 85.6% | 2.1% |
| Logistic Regression | 82.7% | 2.9% |
| LDA | 86.1% | 1.3% |
| QDA | 84.3% | 0.0% |
| K-NN (With K=3) | 84.3% | 1.3% |

Jonathan Pedoeem

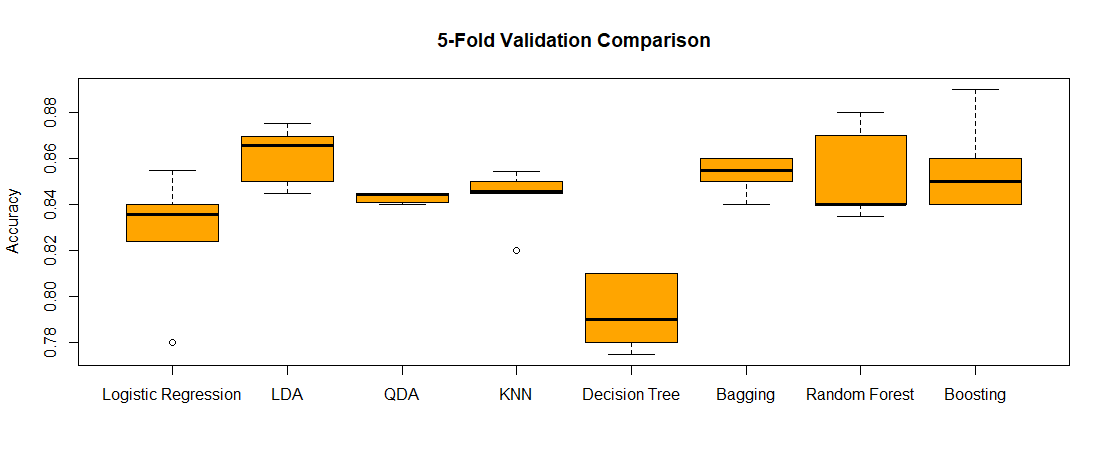
05/31/2018

REU UNCW 2018

Professor Chen

**Project 5**

This project implemented different Tree methods on categorical and numerical data and compared the results using 5-fold cross validation and leave one out cross validation with the four methods from project 4.

*Table 1: Comparison of Gender Classification Using 8 Different Methods with 5-Fold Cross Validation*

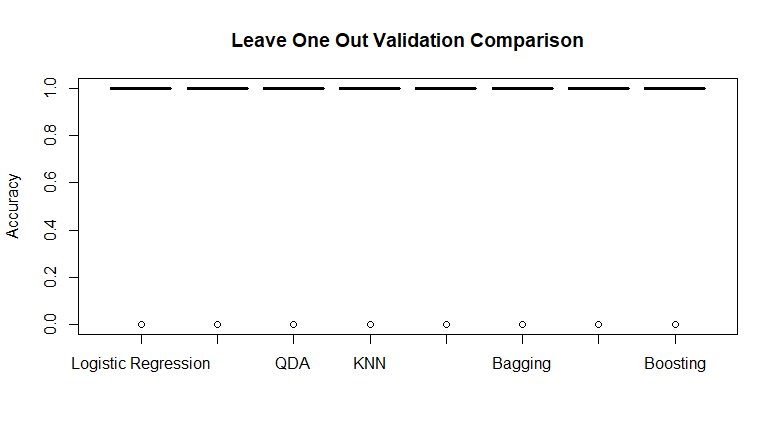
*Figure 1: A Boxplot Comparing all 8 Classification Methods*

The results of the 5-fold CV show LDA having the highest accuracy with Boosting in second place. It is also interesting to note that Random Forest did not have much of an improvement over Bagging.

|  |  |  |  |
| --- | --- | --- | --- |
| **Leave One Out Cross Validation** | | | |
|  | Accuracy | Standard Deviation | Time Taken |
| Decision Tree | 79.8% | 0.402 | 00:03:39 |
| Bagging | 84.3% | 0.364 | 03:28:43 |
| Random Forest | 85.1% | 0.356 | 01:54:40 |
| Boosting | 85.7% | 0.350 | 01:12:44 |
| Logistic Regression | 83.9% | 0.368 | — |
| LDA | 86.2% | 0.345 | — |
| QDA | 84.3% | 0.364 | — |
| K-NN (With K=3) | 84.4% | 0.363 | — |

*Table 2: Comparison of Gender Classification Using 8 Different Methods with Leave One Out Cross Validation*

The results of LOO-CV agree with the rankings of the 5-fold CV. One interesting difference is that Random Forest does have a small improvement over Bagging.



|  |  |  |
| --- | --- | --- |
| **Regression on Morph-II Age Prediction with 5-Fold Cross Validation** | | |
| **Method** | **MAE** | **Standard Deviation** |
| Decision Tree | 6.10 | 0.27 |
| Bagging | 5.59 | 0.25 |
| Random Forest | 5.77 | 0.24 |
| Boosting | 5.48 | 0.32 |

*Table 3: Comparison of Age Regression Using Different Tree Methods with 5-Fold Cross Validation*