Exploring Various Techniques in Surprised Learning

Kam Wing Chung (Kenneth)

Department of Computer Science

Georgia Tech University

Kchung42@gatech.edu

# Purpose

The purpose of this assignment is to explore decision trees, neural networks, boosting, support vector machines and k-nearest neighbor algorithms in supervised learning. All of above supervising learning techniques will be based on Wine Quality Data Set found on (<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>). All of above supervising learning techniques will be using R and R supervising learning packages. R code, dataset and README.txt can be found the same package as this document.

# Wine Quality Dataset and Goal

The two datasets are related to red and white variants of the Portuguese “Vinho Verde” wine. The inputs include object tests and the output is based on sensory data (median of at least three evaluations made by wine experts). Each expert graded the wine quality between 0 (very bad) and 10 (very excellent).

Input variables (based on physiochemical tests and in real values)

1 - Fixed acidity

2 - Volatile Acidity

3 – Citric Acid

4 – Residual sugar

5 – Chlorides

6 – Free sulfur Dioxide

7 – Density

9 – pH

10 – Sulphates

11 – Alcohol

Output variables (based on sensory data)

12 – Quality (score between 0 and 10)

There are no missing attributes values.

There are 4898 white wine samples and 1599 red wine samples.

Here is the distribution of red wine and white wine quality.

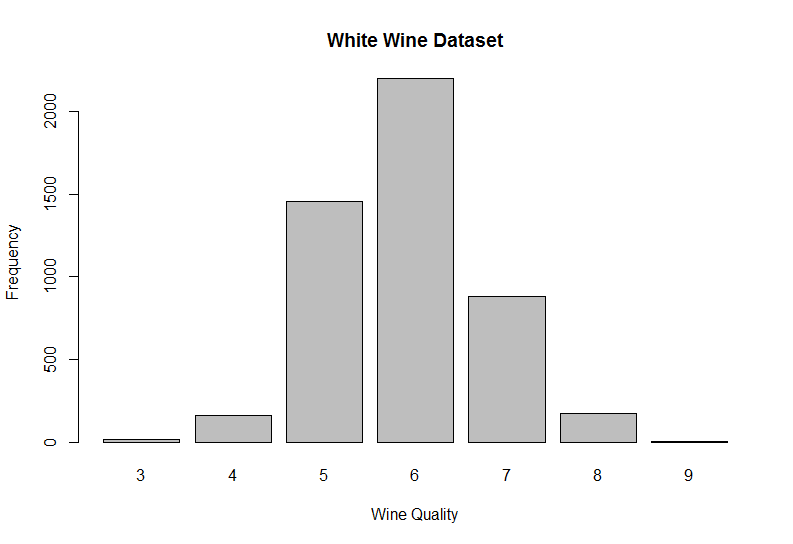


Figure 1

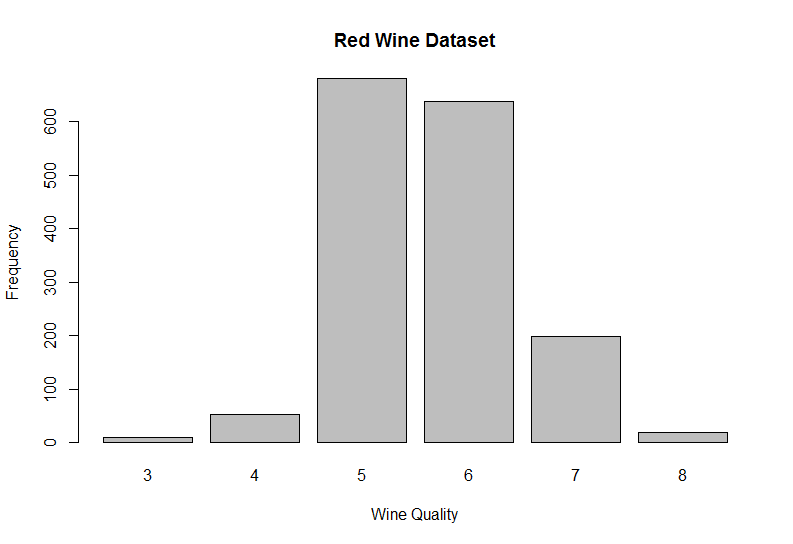


Figure 2

Both white and red wine datasets are divided into half for training and remaining half for testing with random selection from the sample and consistent seed to guarantee same test and training sets for every run.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Quality | White Wine Training | White Wine Test | Red Wine Training | Red Wind Test |
| 1 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 |
| 3 | 10 | 10 | 4 | 4 |
| 4 | 97 | 66 | 21 | 32 |
| 5 | 737 | 720 | 330 | 351 |
| 6 | 1084 | 1114 | 328 | 310 |
| 7 | 430 | 450 | 110 | 89 |
| 8 | 90 | 85 | 6 | 12 |
| 9 | 1 | 4 | 0 | 0 |

Table 1

Table 1 shows the wine quality distribution in four data sets. There are 4898 white wine samples and 1599 red wine samples. Since white has large of size of sample, the following analysis will primary be focusing on white wine dataset.

Since quality attribute values are in numeric and no quality falls between two integers and this is classification problem, the quality attribute values have been converted to characters to fit this goal of this assignment.

The goal for this dataset is classify the test dataset wine quality based on the eleven attributes.

# Decision Tree

## 3.1 No pruning and Analysis

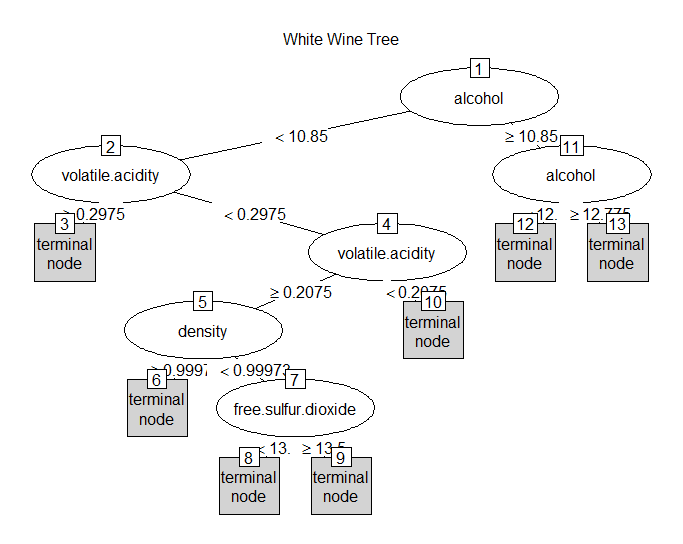


Figure 3

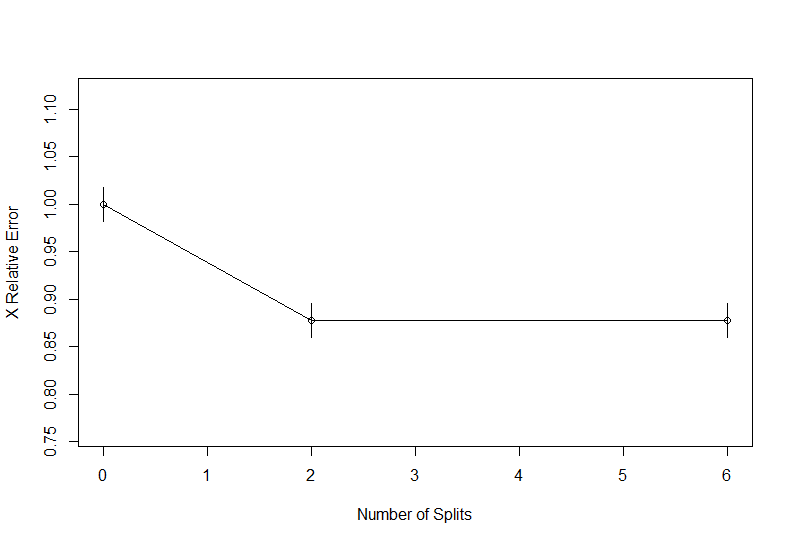


Figure 4

Figure 3 shows that there are six splits on decision tree using the white wine test data. Figure 4 indicates that the starting from split 2, the cross errors remains at just above 0.85.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Predicted Class | | | | | | |
|  |  | 3 | 4 | 5 | *6* | 7 | 8 | 9 |
| Actual Class | 3 | 0 | 0 | 3 | 7 | 0 | 0 | 0 |
| 4 | 0 | 0 | 28 | 38 | 0 | 0 | 0 |
| 5 | 0 | 0 | 377 | 343 | 0 | 0 | 0 |
| 6 | 0 | 0 | 220 | 894 | 0 | 0 | 0 |
| 7 | 0 | 0 | 14 | 436 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 85 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 4 | 0 | 0 | 0 |

Table 2

The probability of predicting correct classes on white wine test dataset is 0.518. 1/n where n is the number of test dataset sample and k is the index of prediction and actual classes. If prediction value equals to actual class, returns 1 otherwise 0.

## 3.2 Pruned Tree and Analysis

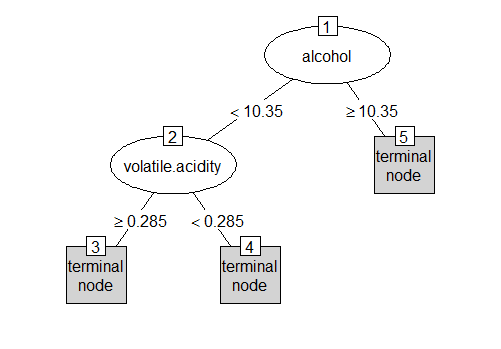


Figure 5

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Predicted Class | | | | | | |
|  |  | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Actual Class | 3 | 0 | 0 | 2 | 8 | 0 | 0 | 0 |
| 4 | 0 | 0 | 21 | 45 | 0 | 0 | 0 |
| 5 | 0 | 0 | 268 | 452 | 0 | 0 | 0 |
| 6 | 0 | 0 | 142 | 972 | 0 | 0 | 0 |
| 7 | 0 | 0 | 10 | 440 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 85 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 4 | 0 | 0 | 0 |

Table 3

Figure 5 has pruned tree based on 2 splits indicated on Figure 4 cross error. The probability of predicting correct class on white wine test dataset using pruned is 0.506 which is slightly lower than non-pruned.

## 3.3 Decision Tree Conclusion

|  |  |  |
| --- | --- | --- |
|  | No Prune | Pruned |
| Probability correct prediction | 0.518 | 0.506 |
| Classes capable of recognizing | 5 and 6 | 5 and 6 |
| Classes incapable of recognizing | 3,4,7 and 8 | 3,4,7 and 8 |
| Number of Splits | 6 | 2 |
| Variables actually used in tree construction | Alcohol, volatile. Acidity, Density, Free Sulfur dioxide | Alcohol and volatile.acidity |
| Stop at Cross Error | ~0.85 | Slightly lower than ~0.85 |

Table 4

The probability of correct prediction has slight dropped in the pruned tree, essentially it has no significant in this just under 2500 samples in both training and test datasets.

Both trees are only capable of predicting class 5 and 6; and incapable of predicting class 3, 4, 7, 8 and 9.

In the pruned tree, there are two out eleven variables are actually used to construct the decision tree.

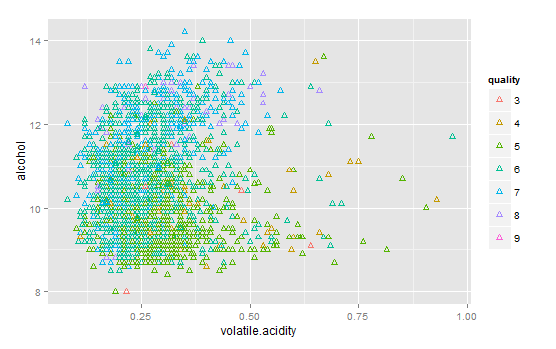


Figure 6

In the figure 6, class 5 wine is populated around low volatile acidity and low alcohol. Class 6 wine is populated around also low volatile and high alcohol. Very small number of class 3 and 8 scattered across. There might be slight greater number of class 4, 7 and 8 that are populated inside the clusters of class 5 and 6.

# Neural Network

Remarks: Due to my computer hardware configuration and performance, the size of training and testing dataset has been scaled down to 1/25 of the size of white wine training and testing data used in above Decision Tree experiment and there are only three attributes (volatile acidity, density and alcohol) being used.

## One Layers and Two Activations

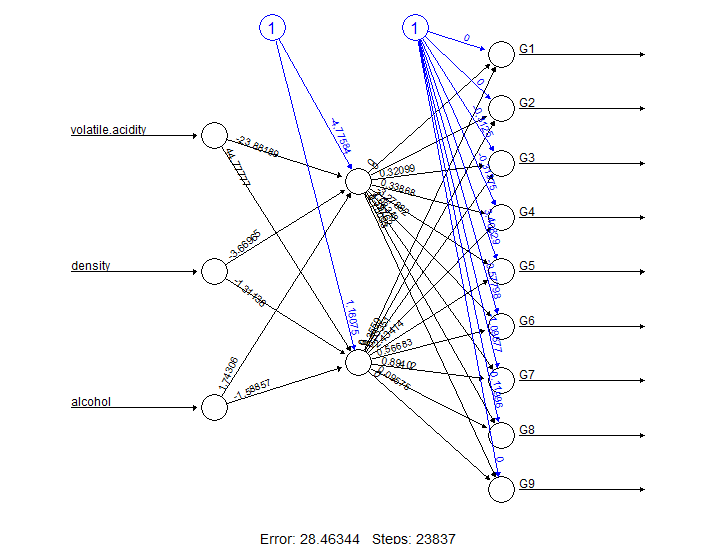


Figure 7

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Prediction Class | | | |
|  |  | 4 | 5 | 6 | 7 |
| Actual Class | 3 | 0 | 1 | 0 | 0 |
| 4 | 0 | 1 | 1 | 0 |
| 5 | 0 | 20 | 11 | 0 |
| 6 | 0 | 3 | 41 | 0 |
| 7 | 1 | 1 | 11 | 0 |
| 8 | 0 | 0 | 4 | 1 |
| 9 | 0 | 0 | 1 | 0 |

Table 5

The probability of predicting correct classes on white wine test dataset is 0.6288 based on about 100 training and test datasets. Even with smaller dataset, the accuracy is higher than Decision Tree with a lot larger sample size.

|  |  |
| --- | --- |
| error | 28.46343968 |
| reached.threshold | 0.009778691 |
| steps | 23837 |

Table 6

In this single layer and two activation, the training process needed 23873 steps until all absolute partial derivatives of error function were smaller than the default threshold 0.01. Since there is only one layer and two activations, the error remains undesirable for solving real problem.

## 2.2 Two Layers and Three Activations

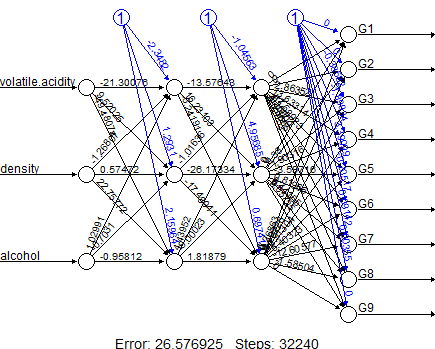


Figure 8

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Prediction Class | | |
|  |  | 4 | 5 | 6 |
| Actual Class | 3 | 0 | 1 | 0 |
| 4 | 0 | 1 | 1 |
| 5 | 1 | 19 | 11 |
| 6 | 1 | 3 | 40 |
| 7 | 2 | 1 | 10 |
| 8 | 0 | 1 | 4 |
| 9 | 0 | 0 | 1 |

Table 7

The probability of predicting correct classes on white wine test dataset is 0.6088 based on about 100 training and test datasets and with two layers and three activations.

|  |  |
| --- | --- |
| error | 26.57693 |
| reached.threshold | 0.009687 |
| steps | 32240 |

The results with additional layer and activation did not improve the accuracy. The reach.threshold has reached lower than default 0.0.1 and the error is 26.57693 slightly lower than single layer error 28.46343968. It could be approaching over fitting. But it is too hard to draw conclusion on these two runs with very small dataset and not enough attributes to try various scenarios.

# Boosting

# Support Vector Machine

# k-nearest neighbors