Machine Leaning Exam Project

Introduction

In this project we will explore how different factors affect wine quality.
We will be using Machine Learning in Python in a Jupyter Notebook and we will use a public dataset.

The dataset

For the dataset, we will be using a famous public dataset that compares different wines both white and red. The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine and we will explore both and see how the modal may differn between them.

The dataset has the following features:
Fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol and quality.

These features are all measured with decimal numbers except for the Quality which is a number from 3 to 8

The model

For our learning method, we will be using RandomForestClassifier, which is a ensemble learning method for classification. We are using Classifier because our Dataset measures quality in a scale from 3 to 8 and does not contain decimal numbers.
By using Classifier instead of Regressor, the algorithm will get the mayority vote from all the decision trees insetead of making an from the results.

Conclusion

When we run the test data trough the model, we were able to get 68-72% acuracy on the Red Wine dataset and a 60-65% acuracy on the White Wine dataset using `n_estimators=1000` (decision trees) and a `max_depth=10` .

We can see the importance of each feature like so:
`array([0.0787448 , 0.1015349 , 0.07661717, 0.07192529, 0.07954605, 0.06568526, 0.09883257, 0.09151635, 0.07962589, 0.1090358 , 0.14693593])`
(In this order: Fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol and quality.)

Finally, I made a mock dataset with some random numbers in a range to give to the train model and receive a predicted quality for set data, which I belive would be the purpose of training this model: to get the predicted quality rating of a new wine.