Wine Quality Detection

Torta Francesco, Voto Giorgio

**Model for the Wine quality detection problem**

*Torta Francesco, Voto Giorgio*Politecnico di Torino, Italy

1. **Introduction**

In this report we will discuss different models able to determine, with a good grade of accuracy, the quality of wine. The original dataset is taken from UCI repository, but has been modified to transform the task into a binary classification problem, with the objetice to label the wine as good or bad in terms of quality.

1. **Feature Analysis**

In the dataset there are 11 features:

1. Fixed acidity
2. Volatile acidity
3. Citric acid
4. Residual sugar
5. Chlorides
6. Free sulfur dioxide
7. Total sulfur dioxide
8. Density
9. PH
10. Sulphates
11. Alcohol

A first analysis of the characteristics shows that in many cases, the features have an uneven distribution. For this reason we will also analyze a dataset pre-processed by Gaussianization, since, especially for models based on Gaussian, we would obtain suboptimal results. The heat map that display the Pearson Correlation Coefficient shows us that some characteristics are related to each other (e.g. 5-6).



While this is true, the correlation is not strong enough to use a feature size reduction technique such as PCA in oru models. This technique will be used only in Gaussian-Based models for completeness of discussion of the problem. In general we will only use K-Fold cross validation with K = 5 to obtain more robust results.

We will analyze the models with 3 different applications:

The main application will be the first. In the other two cases we will have unbalanced tasks that prioritize one of the two class.