Group project 3: Song Hoa Choi, Matt Edwards, Gina Shellhammer, Kalbi Zongo

draft

**“Predict the blind taster quality score of a wine based on chemical tests”**

**Introduction**

To predict the blind taster quality score of a wine based on chemical tests, we collected “Wine Quality” data from UCI Machine Learning Repository, <https://archive.ics.uci.edu/ml/datasets.html>. There are twelve variables such as fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, ph, sulphates, alcohol, and taster quality scale 0 to 10 in two datasets, 4898 white and 1599 red vinho verde wine samples from Northern Portugal. The taster quality scale 0 indicates ‘very bad’ and 10 indicate ‘very excellent’. The response variable is the taster quality scale with eleven explanatory variables from various phytochemicals in wine. The median of taster quality in white (n = 2198) and red wine (n = 681) is between 5 and 6. We decided to use 5 and 6 from taster quality scale to predict.

**Description of the machine learning method**

There are steps to obtain the finding:

1. Stratified sampling method

2. Principal component analysis (?)

3. Ordinal regression

First, we constructed training and testing set by using stratified the quality variables. 37.5% of items in strata were randomly selected to be in the testing set and remaining 62.5% were the training set. Secondly, principal component analysis was conduct. Finally, ordinal regression---I will write down in detail..depending on our project progress..

-------------------

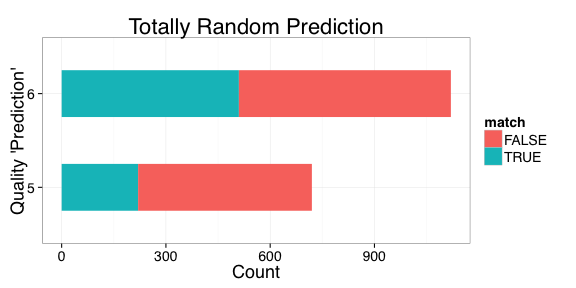
**Summary findings**

**-----------**

**Discussion including assumptions/limitations**

**-------------**

**Appendix**

****