

Project Proposal

-----Classifying Wine Quality

This project is about exploring the performance of different algorithms on the same dataset. The dataset in the project is called “*winequality-white*” sourced from <https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/>.

This dataset has 12 features and a total of 4898 entries. The total number of data points is 58776. The dimension of this dataset is 4898 by 12. Labels or features include fixed acidity; volatile acidity; citric acid; residual sugar; chlorides; free sulfur dioxide; total sulfur dioxide; density; pH; sulphates; alcohol; quality. These 12 features will be reduced to a smaller number if necessary during experiments.

The classification problem that will be explored on this dataset is to classify the wine as a “good” wine if the quality number greater than 6.5.

The first algorithm that will be explored in this project is linear regression which is the most basic classification method. The second algorithm will be K-means clustering, and the third algorithm to explore is neural networks.

Some parameters that will be explored for these three classification algorithms include:

1. The degree of the polynomial in linear regression.
2. Different sets of features in K-means clustering.
3. Different layers of neural networks.

The entire dataset will be divided into a training set and a validation set. The performance on the validation set after training will be used to compare these algorithms.

The project is hosted on GitHub and here is the link:
<https://github.com/guang16/Machine-Learning>

A brief timeline:

10/22: Initial project proposal

10/23 – 11/6: Explore linear regression and potentially modify the data by extracting some features.

11/7 – 11/17: Explore K-means clustering, adjust timeline and construct project report 1

11/17: Project report 1

11/18 –12/1: Explore Neural Networks, finishes up with the first 2 algorithms if necessary, and construct project report 2.

12/1: Project report 2

12/1 – 12/12: Construct the final report

12/12: Final project report