

CITS3401 Data Exploration and Mining Project 2

Wine Classification

Mitchell Pomery
21130887

June 1, 2014

Introduction

The project specified that we are to compare how machine learning performs compared to experts when rating different wines. The initial data is split into two groups, red wine and white wine. The dataset is available from the UCI Machine Learning Repository[1].

Data analysis is done using Weka[2], data mining software created by Machine Learning Group at the University of Waikato.

Data Preprocessing

The initial data provided was in two files, `winequality-red.csv` and `winequality-white.csv`, that were converted to Weka's ARFF file format using an online conversion tool. This tool was used to output two datasets, dataset 1 (`ds1-red.arff` and `ds1-white.arff`) and dataset 2 (`ds2-red.arff` and `ds2-white.arff`). Dataset 1 contains all the information that was in the original data, and is used to create the classifier. The fields in this dataset are numeric, apart from the quality which is nominal, making it possible to group wines that receive the same rankings in Weka.

Dataset 2 contains all the numerical information from the original data and does not contain any information about the rankings from the wine tasters.

Classification

Support Vector Machine

[4]

Neural Network

Naive Bayesian

Results

Bibliography

- [1] UCI Machine Learning Repository: Wine Quality Data Set. 2014. UCI Machine Learning Repository: Wine Quality Data Set. [ONLINE] Available at: <http://archive.ics.uci.edu/ml/datasets/Wine+Quality>. [Accessed 01 June 2014].
- [2] Weka 3 - Data Mining with Open Source Machine Learning Software in Java . 2014. Weka 3 - Data Mining with Open Source Machine Learning Software in Java . [ONLINE] Available at: <http://www.cs.waikato.ac.nz/ml/weka/>. [Accessed 01 June 2014].
- [3] Online CSV to ARFF conversion tool. 2014. Online CSV to ARFF conversion tool. [ONLINE] Available at: <http://slavnik.fe.uni-lj.si/markot/csv2arff/csv2arff.php>. [Accessed 01 June 2014].
- [4] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.