CKMT 105
Homework 3
Deadline: 20 April, 2015

PROBLEM DEFINITION

Build a model to predict the quality of red wine using its attributes..

This is an informally defined ptoblem. First formally define this problem and then build a model and report your findings.

DATA

The dataset is related to white Portuguese "Vinho Verde" wine. For more info go: https://archive.ics.uci.edu/ml/datasets/Wine+Quality

TASK

1. (10 points) Import the data to R (no cleanup necessary).
2. (10 points) Check data characteristics. Is there missing data? What is the correlation?
3. (30 points) Propose a model for the prediction. Give a few reasons for your selection briefly. You may choose to model the problem as classification or regression. Define the task, experience and performance criteria.
4. (30 points) Do 10 fold cross validation in the experiment.
5. (20 points) Report your results.
6. (Bonus 20 points) Other researchers have built predictive models using this data. Do a brief comparison of your results with the result in the following paper: http://repositorium.sdum.uminho.pt/bitstream/1822/10029/1/wine5.pdf

OUTPUT

A zip file containing the following
- A report with:
    - Model
    - Results
    - Brief discussion of the work done for each task
- All the source code