ASSIGNMENT 3
2016 FALL TERM

CMTH 642 Data Analytics: Advanced Methods
Course Instructor: Ceni Babaoglu

Preparation:

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

Import to R the following file:
http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv

TASK

1. Check data characteristics. Is there missing data?
2. What is the correlation between the attributes other than wine quality?
3. Graph the distribution of wine quality.
4. Reduce the levels of rating for quality to three levels as high, medium and low.
5. Normalize the data set.
6. Divide the data to training and testing groups.
7. Use the KNN algorithm to predict the quality of wine using its attributes.
8. Evaluate the model performance.

Save your R codes in an RMD file.
Send your file to the course shell.