

## Assignment #4: Predictive Modeling in Multiclass Classification

**Data:** The data for this assignment is the 'Wine' data set. The data must be downloaded from the UC Irvine Machine Learning Repository. The data will need to be read into R using the function `read.csv()`.

<http://archive.ics.uci.edu/ml/datasets/Wine>

### Assignment Instructions:

This assignment will build upon the previous work completed in Assignment #1. Use some of the output from the analysis completed in Assignment #1 to fill out a report consistent with the quality required by Predict 454. In this assignment we will focus on adding the modeling component to the analysis of the wine data. Since this is a small data set we will only focus on fitting models in-sample.

#### (1) Fit a Model Suite:

For this data we will fit a model suite of the following models: (1) Random Forest, (2) a Support Vector Machine, and (3) a neural network model.

- Note that this is a multiclass classification problem. How do we compare and contrast predictive accuracy in a multiclass classification problem? Hint: ROC curves are not valid in a multiclass classification problem.
- Note that these are typical 'black box' models. Do any of these models allow any type of model visualizations? Can we produce any type of output that can help us 'understand' the models?
- Since these are 'black box' models, how should we communicate the results of these models?
- Use tables when needed. Do not simply cut and paste R output.

### Presentation:

The presentation of your results is open to you. Remember that the objective of your analysis is to effectively **communicate the important information**. At a minimum your presentation should include a section for the data quality check (where you discuss the overall approach that you used and any interesting results), a section for the exploratory data analysis (where you present and discuss your interesting findings, and an appendix of relevant R code related to the results or graphics that you have presented.

**Assignment Document:**

Students should present their results in the form of a report. Reports should be well written. Results should be embedded into the report in the sections with the corresponding discussion. All figures and tables should be centered and labelled. Samples of relevant R code related to the output discussed in the report should be included in an appendix. This should not be all of your R code, only some of the relevant R code. For example if you made a very specific statistical graphic that you discuss in the report, then the R code for that statistical graphic should be included in your appendix. The R code used to fit any statistical models should always be included in the appendix.

The report document should be submitted in pdf format.