HW #1
Due: November 29, 2019

Student ID: _____ Name: _____

The `wine.csv` data is the results of a chemical analysis of wines grown in Italy. The analysis determined the quantities of 9 constituents found in each of two types of wines. In this home assignment, we build classification models based on this dataset. For each of the following questions, show all of your works in detail.

1. Building a decision tree
   - ① Build a decision tree using `rpart()` command. Show your R code for reading the given dataset and creating an R object which is the decision tree for classifying types of wines.
   - ② Plot your decision tree and show the classification rule on it.
   - ③ Which variables are the most important to classify the types of wines?
   - ④ Compute the training error (misclassification rate) of your decision tree. Show your R code.
   - ⑤ Apply the `wine_test.csv` data into your decision tree in order to predict the type of wines. Show your R code and report the predicted class labels.
   - ⑥ Based on your answer to question #3, select two most important variables and plot the wine data (training set) on the two-dimensional space of the selected variables. Distinct types of wines by coloring the observations with two colors.

2. Comparing classification methods
   - ① Based on the `wine.csv` data, build classification models using logistic regression, discriminant analysis, k-nearest neighbors, neural networks, and naïve Bayes. Compare the classification performances of the trained classifiers by reporting their training errors and test errors (the test set is `wine_test.csv`.). Show your R code.
   - ② What is your best performing classifier?

✓ Your homework must be turned in as a hard copy before the class begins. Please make sure that NO late homework is accepted.