Homework 4

STAT 430, Spring 2017

Due: Friday, February 24 by 11:59 PM

Please see the homework instructions document for detailed instructions and some grading notes. Failure to follow instructions will result in point reductions.

Exercise 1

[10 points] We once again return to the data found in `auto-train.csv` and `auto-test.csv` which contain train and test data respectively. `auto.csv` is provided but not used. It is a modification of the `Auto` data from the `ISLR` package. Use `mpg` as the response. Train the following models:

- Additive Logistic Regression
- LDA
- QDA
- Naive Bayes

Report test and train accuracies for both. Indicate which model performs best.

You should consider coercing the response to be a factor variable.

Exercise 2

[20 points] Use the data found in `hw04-train.csv` and `hw04-test.csv` which contain train and test data respectively. Use `y` as the response. Coerce `y` to be a factor after importing the data.

Create an ellipse plot then train the following models:

- Additive Logistic Regression
- LDA
- LDA with Flat Prior
- QDA
- QDA with Flat Prior
- Naive Bayes

Report test and train accuracies for both. Indicate which model performs best. Do the results match the intution from the plot? Which class(es) is your best classifier classifying the best?