Homework 3

STAT 430, Spring 2017

Due: Friday, February 17 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.

For this homework we 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.

We will use this data for each exercise in this homework.

For information on the original data:

```
library(ISLR)
#?Auto
```

Exercise 1

[10 points] Use the training data to fit both a linear and logistic regression using only `displacement` as the predictor. Use both to create classifiers which seek to minimize the classification error.

For both:

- Plot the training data and add a line (or curve) with the predicted probabilities.
- Find decision boundary c. That is, find c such that

$$\hat{C}(\text{displacement}) = \begin{cases} 0 & \text{displacement} > c \\ 1 & \text{displacement} \leq c \end{cases}$$

- Report the test accuracy.

Exercise 2

[12 points] Now consider a logistic regression that considers two predictors, `acceleration` and `weight` in an additive model. Do the following:

- Plot the training data with `acceleration` as the x axis, and `weight` as the y axis, with the points colored according to their class. Add a line which represents the decision boundary for a classifier using 0.5 as a cutoff for predicted probability. **This may be challenging.**
- Report test sensitivity, test specificity, and test accuracy for three classifiers, each using a different cutoff for predicted probability:
  - 0.2
  - 0.5
  - 0.8
- Plot an ROC curve and report the AUC.

Exercise 3

[8 points] Finally, consider the full additive logistic regression. Create an improved model for classification by adding (or removing) complexity. Report relevant metrics for both models to justify your model.