STATS 415 Homework 5

Due Thursday Feb 15, 2018

Please include your name, uniqname, and lab section (number or time or GSI). A point will be taken off homework without the section info. Turn in a printout of your homework in the lecture or in your GSI's mailbox across room 305A West Hall, no later than 5pm on the due date.

This homework continues homework 4. Use the same data, the same split into training and test, and the same four variables you chose to use as predictors.

1. Perform logistic regression on the training data in order to predict `mpg01` using the four quantitative variables you chose in Homework 4. Comment on the significance of the coefficients.

2. Report the training and the test errors for logistic regression. Make a plot similar to the plots in HW4, showing true and predicted class labels from logistic regression plotted against the same two variables you used before.

3. Using your fitted model, estimate the probability of a car having above-median mpg if its four predictors you used are all at the median values for the training dataset.

4. Perform KNN classification on the training data. Make plots of the training classification error and the test classification error as a function of the number of neighbors K (or $1/K$; if you use $1/K$, make sure the x-axis is on the log scale). Which K gives the best performance on the training data? On the test data?

5. Report the training and the test errors for KNN with your choice of K. Make a plot similar to the plots in HW4, showing true and predicted class labels from KNN plotted against the same two variables you used before.

6. Can you answer question 3 with KNN regression? If yes, give the answer. If not, explain why not and what you can report instead for a car with the four predictors all at the median values.

7. Compare and contrast the performance of LDA, QDA (take from HW 4), logistic regression, and KNN on this dataset. What do your results suggest about the distribution of the data? About the nature of the boundary between classes?

Please limit your solution to at most 6 pages.