

Homework 3

Due on 04/09/2019

This questions will be answered using the `Weekly` data set, which is part of the `ISLR` package. This data is similar in nature to the `Smarket` data on the textbook except that it contains 1,089 weekly returns for 21 years, from the beginning of 1990 to the end of 2010. A description of the data can be found by typing `?Weekly` in the Console. (Note that the column `Today` is not a predictor.)

- (a) Produce some graphical summaries of the `Weekly` data.
- (b) Use the full data set to perform a logistic regression with `Direction` as the response and the five `Lag` variables plus `Volume` as predictors. Do any of the predictors appear to be statistically significant? If so, which ones?
- (c) Compute the confusion matrix and overall fraction of correct predictions. Briefly explain what the confusion matrix is telling you.
- (d) Plot the ROC curve using the predicted probability from logistic regression and report the AUC.
- (e) Now fit the logistic regression model using a training data period from 1990 to 2008, with `Lag1` and `Lag2` as the predictors. Plot the ROC curve using the held out data (that is, the data from 2009 and 2010) and report the AUC.
- (f) Repeat (e) using LDA and QDA.
- (g) Repeat (e) using KNN. Briefly discuss your results.