Weekly Exercise - Week 3

1 Logistic Regression Computations

Complete Exercise 6 from the book on page 191.

2 Logistic Regression Applied

Answer the following, easier variant of Exercise 13 from the book on page 193, using the `Weekly` dataset, which is part of the `ISLR2` package. This dataset is similar to the `Smarket` data from this chapter's lab, except that it contains 1,089 weekly returns for 21 years, from the beginning of 1990 to the end of 2010. If you need a reminder of the data set have a look at `?Weekly`.

- Fit a logistic regression model using a training data period from 1990 to 2008, with Direction as the response and the five lag variables plus Volume as predictors.
- Use the summary function to print the results. Do any of the predictors appear to be statistically significant? If so, which ones?
- Use the held-out data (that is, the data from 2009 and 2010) to compute the balanced accuracy. Would you trust the model?

3 Comparison of Methods

- Which method has higher variance: LDA or logistic regression? What about bias?
- Assuming LDA's assumptions are perfectly met, which method will perform better: LDA or logistic regression?
- Assuming the decision boundary is very nonlinear, the number of features is $p = 1$, and the training set size is $n = 1,000,000$, which of the following methods do you expect to perform best: logistic regression, LDA, QDA, kNN?