WEEKLY STOCK MARKET DATA

This question uses the *Weekly* dataset in the *ISLP*. It contains 1089 weekly stock returns for the 21 years between the beginning of 1990 to the end of 2010. You can access the data using in Python

$$Sweekly = load\_data('Weekly')$$

(a) Produce some numerical and graphical summaries of the Weekly data. Are there any apparent patterns?

(b) Use the full data to perform a logistic regression with Direction as the response variable and the five lags variables plus Volume as predictors. Do any of the predictors appear to be statistical significant? If so which ones?

(c) Compute the confusion matrix and overall fraction of correct predictions. Explain what the confusion matrix is telling you about the types of mistakes logistic regression is making.

(d) Now fit the logistic regression using a training data period from 1990 to 2008 and Lag2 as the only predictor. Compute the confusion matrix and overall fraction of correct predictions for the hold out data, i.e., 2009 and 2010.

(e) Repeat (d) using the linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA).

(f) For the test data using KNN, plot the misclassification error rate vs $1/k$. What is the optimal k that minimizes the test misclassification error rate?

(g) Which of these various methods appears to provide the best results on this data?

(h) Plot the ROC curves for different classifiers, e.g. logistic regression, LDA, KNN with different k values and discuss the performance (the larger the area under the curve, the better the classifier).