Homework 0408

1. This question should be answered using the *Weekly* data set, which is part of the ISLR2 package. This data contains 1089 weekly returns for 21 years, from the beginning of 1990 to the end of 2010.
   (a) Produce some numerical or graphical summaries of the Weekly data. Do there appear to be any patterns? (use *pairs( ), cor( )*)
   (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. Use the summary function to print the results. Do any of the predictors appear to be statistically significant? If so, which ones? Give a numerical interpretation to the effect of the significant predictors.
   (c) Compute the confusion matrix and overall fraction of correct predictions. Explain what the confusion matrix is telling you.
   (d) Now fit the logistic regression model using a training data period from 1990 to 2008, with Lag2 as the only predictor. Compute the confusion matrix and the overall fraction of correct predictions for the held out data (That is, the data from 2009 and 2010).

2. Let $x = (x_1, x_2)'$. Prove that

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

is equivalent to

$$p(x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}.$$