HW9

Math 189

Friday of Week 9, 05/27/2022

1. This question uses the variables `dis` (the weighted mean of distances to five Boston employment centers) and `nox` (nitrogen oxides concentration in parts per 10 million) from the `Boston` data. We will treat `dis` as the predictor and `nox` as the response.

   (a) Use the `bs()` function to fit a spline regression to predict `nox` using `dis`. Report the regression output, and plot the resulting data and the fitted curve.

   (b) Use the `ns()` function to fit a regression spline to predict `nox` using `dis`. Report the output for the fit using **four** degrees of freedom. How did you choose the knots? Plot the resulting fit.

   (c) Now fit a regression spline for a range of degrees of freedom, and plot the resulting fits and report the resulting RSS. Describe the results obtained.

   (d) Perform cross-validation in order to select the best degrees of freedom for a regression spline on this data. Describe your results.

   (e) Use the `smooth.spline` function to fit a smooth spline regression to predict `nox` using `dis`. Choose the best `df` by using LOOCV. Report your best `df` and plot the results.

2. In this question, you are asked compute the LOOCV error for a simple logistic regression model on the `Weekly` data set.

   (a) Fit a logistic regression model that predicts `Direction` using `Lag1` and `Lag2`.
   (b) Fit a logistic regression model that predicts `Direction` using `Lag1` and `Lag2` using all but the first observation.
   (c) Use the model from (b) to predict the direction of the first observation. You can do this by predicting that the first observation will go up if $P(\texttt{Direction="Up"}|\texttt{Lag1,Lag2}) > 0.5$. Was this observation correctly classified?
   (d) Write a "for loop" from $i = 1$ to $i = n$, where n is the number of observations in the data set, that performs each of the following steps:
      i. Fit a logistic regression model using all but the ith observation to predict `Direction` using `Lag1` and `Lag2`.
      ii. Compute the posterior probability of the market moving up for the ith observation.
      iii. Use the posterior probability for the ith observation in order to predict whether or not the market moves up.
      iv. Determine whether or not an error was made in predicting the direction for the ith observation. If an error was made, then indicate this as a 1, and otherwise indicate it as a 0.
   (e) Take the average of the n numbers obtained in (d)iv in order to obtain the LOOCV estimate for the test error. Comment on the results.