Assignment #2: Linear Regression

Problem 1

Use the Auto data set to answer the following questions:

(a) Perform a simple linear regression with mpg as the response and horsepower as the predictor. Comment on the output. For example

i. Is there a relationship between the predictor and the response?
ii. How strong is the relationship between the predictor and the response?
iii. Is the relationship between the predictor and the response positive or negative?
iv. How to interpret the estimate of the slope?
v. What is the predicted mpg associated with a horsepower of 98. What are the associated 95% confidence and prediction intervals?

(b) Plot the response and the predictor. Display the least squares regression line in the plot.

(c) Produce the diagnostic plots of the least squares regression fit. Comment on each plot.

(d) Try a few different transformations of the predictor, such as $\log(X)$, $\sqrt{X}$, X^2, and repeat (a)-(c). Comment on your findings.

Problem 2

Use the Auto data set to answer the following questions:

(a) Produce a scatterplot matrix which includes all of the variables in the data set. Which predictors appear to have an association with the response?

(b) Compute the matrix of correlations between the variables (using the function cor()). You will need to exclude the name variable, which is qualitative.

(c) Perform a multiple linear regression with mpg as the response and all other variables except name as the predictors. Comment on the output. For example,

i. Is there a relationship between the predictors and the response?
ii. Which predictors have a statistically significant relationship to the response?
iii. What does the coefficient for the year variable suggest?

(d) Produce diagnostic plots of the linear regression fit. Comment on each plot.

(e) Is there serious collinearity problem in the model? Which predictors are collinear?

(f) Fit linear regression models with interactions. Are any interactions statistically significant?

Problem 3

Use the Carseats data set to answer the following questions:

(a) Fit a multiple regression model to predict Sales using Price, Urban, and US.

(b) Provide an interpretation of each coefficient in the model (note: some of the variables are qualitative).

(c) Write out the model in equation form.

(d) For which of the predictors can you reject the null hypothesis $H_0: \beta_j = 0$?

(e) On the basis of your answer to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the response.

(f) How well do the models in (a) and (e) fit the data?

(g) Is there evidence of outliers or high leverage observations in the model from (e)?

Submit through link: eCampus -> Assignments->Assignment 2 Submission

Deadline: Oct 3 (Tue) @11:59pm