

## Lab 11 Assignment

Dr. Purna Gamage

**Problem 1 : Answer all parts except part (d).**

9. This question involves the use of multiple linear regression on the `Auto` data set.
- (a) Produce a scatterplot matrix which includes all of the variables in the data set.
  - (b) Compute the matrix of correlations between the variables using the function `cor()`. You will need to exclude the `name` variable, which is qualitative. `cor()`
  - (c) Use the `lm()` function to perform a multiple linear regression with `mpg` as the response and all other variables except `name` as the predictors. Use the `summary()` function to print the results. Comment on the output. For instance:
    - i. Is there a relationship between the predictors and the response?
    - ii. Which predictors appear to have a statistically significant relationship to the response?
    - iii. What does the coefficient for the `year` variable suggest?
  - (d) Use the `plot()` function to produce diagnostic plots of the linear regression fit. Comment on any problems you see with the fit. Do the residual plots suggest any unusually large outliers? Does the leverage plot identify any observations with unusually high leverage?
  - (e) Use the `*` and `:` symbols to fit linear regression models with interaction effects. Do any interactions appear to be statistically significant?
  - (f) Try a few different transformations of the variables, such as  $\log(X)$ ,  $\sqrt{X}$ ,  $X^2$ . Comment on your findings.

```
library(ISLR)
```

```
## Warning: package 'ISLR' was built under R version 4.1.1
```

For part (f): Hint: To determine if nonlinear transformations may be useful, make scatterplots of the response on each quantitative predictor.

## Problem 2: Answer for parts (a)-(g)

10. This question should be answered using the **Carseats** data set.

- (a) Fit a multiple regression model to predict **Sales** using **Price**, **Urban**, and **US**.
- (b) Provide an interpretation of each coefficient in the model. Be careful—some of the variables in the model are qualitative!
- (c) Write out the model in equation form, being careful to handle the qualitative variables properly.
- (d) For which of the predictors can you reject the null hypothesis  $H_0 : \beta_j = 0$ ?
- (e) On the basis of your response to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.
- (f) How well do the models in (a) and (e) fit the data?
- (g) Using the model from (e), obtain 95% confidence intervals for the coefficient(s).
- (h) Is there evidence of outliers or high leverage observations in the model from (e)?

For part (f): Hint: Use  $R^2$  to compare.