STT 811

Homework 3

Due Wednesday, February 8, at 11:59:59 pm

1. **ISLR2** 4.6

2. **IRLR2** 4.9

3. **ISLR2** 4.14
    a. Part (a)
    b. Part (b) (be careful as to which X variables are truly numerical (or not))
    c. Part (c)
    d. Part (f)
    e. Use the parameters found in the logistic regression to compute the predictions directly for the test dataset. Compare to the output of the predict function.
    f. Create a confusion matrix for the train and test datasets. Does the model perform similarly between them? Be sure to explain.
    g. Compute confidence intervals for the parameters based on the z values.
    h. Compute bootstrapped confidence intervals for the parameters you found.
    i. Re-do the logistic regression with 2 numerical variables (no train-test split). Create a contour plot of the prediction results.

4. For the customer churn dataset, consider the fields Age, Total_Purchase, Account_Manager, Years, and Num_Sites as possible X variables. Note that Account_Manager is a binary categorical variable.
    a. Create histograms to examine how each variable might predict churn.
    b. Split the data into train and test datasets.
    c. Fit a logistic regression model—first with all X's, and then remove those X's that are not statistically significant, one at a time. Create a confusion matrix for the final model for both the train and test datasets, and compare the results.