STT 811

Homework 4

Due Wednesday, February 22, at 11:59:59 pm

1. **ISLR2** 4.14 (for all parts do a train-test split like in HW 3 (use the same train-test split for each))
    a. Part (g)
    b. Part (d)
    c. Part (h)
    d. Re-do the naïve Bayes calculation from first principles (i.e. without any packages, by calculating the class means and standard deviations)
    e. Do a modified naïve Bayes model (2 numerical X's) which takes into account the class covariances between the X's.
    f. Create confusion matrices and compute the overall accuracy for the 5 models (test dateset). Compare how the models did.

2. 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. After doing a train/test split,
    a. Create a naïve Bayes model to predict churn.
    b. Create a KNN neighbors model to predict churn. Vary K from 4 to 10 and find out which K has the highest accuracy.
    c. Create the confusion matrices for the test dataset for each of these and compare the models' performance.