Home Work # 6. AMS 380

Name:_____SBU ID:_____

Dear all, the homework is due on Tuesday, Oct 26, 2021, at 11:59 PM. Please submit your homework to the Blackboard in a pdf or word (.doc) document. Rmarkdown is highly recommended.

Please include (1) R code; (2) Output from R; (3) Answers to all the questions asked.

Logistic Regression (A type of Generalized Linear Model) with the Banknote Data*

The banknote.csv data (see attached) were extracted from images that were taken from genuine and forged banknote-like specimens. Yes, this is a *Catch Me if You Can* story. For digitization, an industrial camera usually used for print inspection was used. The final images have 400x 400 pixels. Wavelet Transform tool were used to extract features from images. **There are 1372 banknotes, and 5 variables:**

1. variance of Wavelet Transformed image (continuous)
2. skewness of Wavelet Transformed image (continuous)
3. curtosis of Wavelet Transformed image (continuous)
4. entropy of image (continuous)
5. class (binary) - this is the response variable of interest, 1 (forged) or 0 (genuine).

follow the procedures from the following website:
http://www.sthda.com/english/articles/36-classification-methods-essentials/151-logistic-regression-essentials-in-r/
http://atm.amegroups.com/article/view/9706/pdf

1. (a) Please split the data into 80% training and 20% testing using seed =123.
   (b) Then you shall fit a logistic regression model with all 4 predictors using the training data.
   (c) Please use this fitted model based on the training data to predict the response variable 'class' (whether the banknote is forged or not) for the testing data. Please generate the confusion matrix, and report:
   - (i) The overall accuracy;
   - (ii) The sensitivity (that is, the probability a banknote is predicted to be forged given that it was in fact forged);
   - (iii) The specificity (that is, the probability a banknote is predicted to be genuine given that it was in fact genuine).

1. Please find a model that best predicts whether the banknote is forged or genuine using the stepwise variable selection method and the BIC, based on the entire dataset. Please only use the original variables and do not include any other variables such as interactions. Please report the final model and the associated BIC value.

2. Please find a model that best predicts whether the banknote is forged or genuine using the best subset variable selection method and the BIC, based on the entire dataset. Please only use the original variables and do not include any other variables such as interactions. Please report the final model and the associated BIC value.