Homework #1. AMS 580

Name:_____SBU ID:_____

Dear all, this homework is due by 8am, February 13. Please submit your homework to the Brightspace in two files: (1) the RMD file., and (2) the output of the RMD file – saved as a word document or PDF file. If you have trouble submitting the quiz to the Brightspace, please email it to your TA Ian at: weihao.wang@stonybrook.edu

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).

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).