# Quiz 6. AMS 597

# Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_SBU ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# The quiz is due at the end of the lecture by 11:20am – please submit no later than 11:20am. Please email your completed quiz to your TA at: [song.jiecheng@stonybrook.edu](mailto:song.jiecheng@stonybrook.edu)

# Please include (1) R code; (2) Output from R;

# (3) Answers to all the questions asked

# Please keep yourself on Zoom video until you have emailed your solutions.

# Please plug your computer in power source to avoid running low on battery.

#### Logistic Regression with the Banknote Data – Classification

The banknote 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, 0 (forged) or 1 (genuine).

Your task is to split the data randomly into training (75%) and testing (25%), establish an optimal model using the training data, and then use that model to predict whether each banknote in the testing data is genuine or forged.

1. For the training data, please find a model that best predicts whether a banknote is genuine or forged using the stepwise variable selection method, considering all 2-way interactions, and using the BIC. Please report the final model and the associated BIC value.

1. Please compute the Confusion matrix for the training data – please report the sensitivity, specificity and the overall accuracy based on the training data.
2. Please use the above model to predict whether each banknote in the testing data is genuine or forged. Please compute the Confusion matrix for the testing data based on the default threshold of 0.5 – please report the sensitivity, specificity and the overall accuracy based on the testing data.
3. Please plot the ROC curve for the prediction of the testing data. What is the AUC value of your ROC curve?
4. (Extra credit) Can you perform the prediction using the Jackknife cross-validation (that is, the leave-one-out) method? Please compute the Confusion matrix for the results based on the default threshold of 0.5 – please report the sensitivity, specificity and the overall accuracy.