# Quiz 8. AMS 580

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

# The quiz is due at by 8:00am – on Tuesday, May 4, 2021. Please submit your quiz in one word or pdf file to the Blackboard.

# If any difficulty with the Blackboard, please email it to your TA at: [weihao.wang@stonybrook.edu](mailto:weihao.wang@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 submitted your solutions.

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

#### SVM with the Banknote Data – Classification Task

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 1,372 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).

First, one must clean the data for missing values and deleted observations with missing value(s). Next you need to split the data randomly into training (75%) and testing (25%), first build the best SVM models to predict ‘class’ (whether check is forged or genuine) using the training data, and then use these models to predict whether each check in the testing data is forged or not. Please use the *caret* package in R to build the various SVM classifiers.

**Note:** For this data set, we shall need to perform data standardization for all the continuous variables.

Please review the following website for related methods and concepts:

<http://www.sthda.com/english/articles/36-classification-methods-essentials/144-svm-model-support-vector-machine-essentials/>

1. For the entire dataset, please perform the data cleaning as instructed before; namely, delete observations with missing value(s). Please report how many checks are left after this step. Then please use the random seed 123 to divide the cleaned data into 75% training and 25% testing.

1. Please first build the best classifier to predict whether a check is forged or not using the training data and the linear SVM. We shall use the default value for the cost parameter C. Please compute the Confusion matrix and report the sensitivity (that is, a forged check is found to be forged), specificity (that is, a genuine check is found to be genuine), and the overall accuracy using the testing data.
2. Next we will build the best classifier to predict whether a check is forged or not using the training data and the linear SVM. We shall find the optimal cost parameter C by using the command line:

tuneGrid = expand.grid(C = seq(0, 2, length = 20))

Please (i) report the optimal cost parameter value, and (ii) compute the Confusion matrix and report the sensitivity, specificity, and the overall accuracy using the testing data.

1. Now we shall build the best classifier to predict whether a check is forged or not using the training data and the SVM with radial basis kernel. We shall find the optimal tuning parameters C and sigma (**σ**) by using the command line:

tuneLength = 10

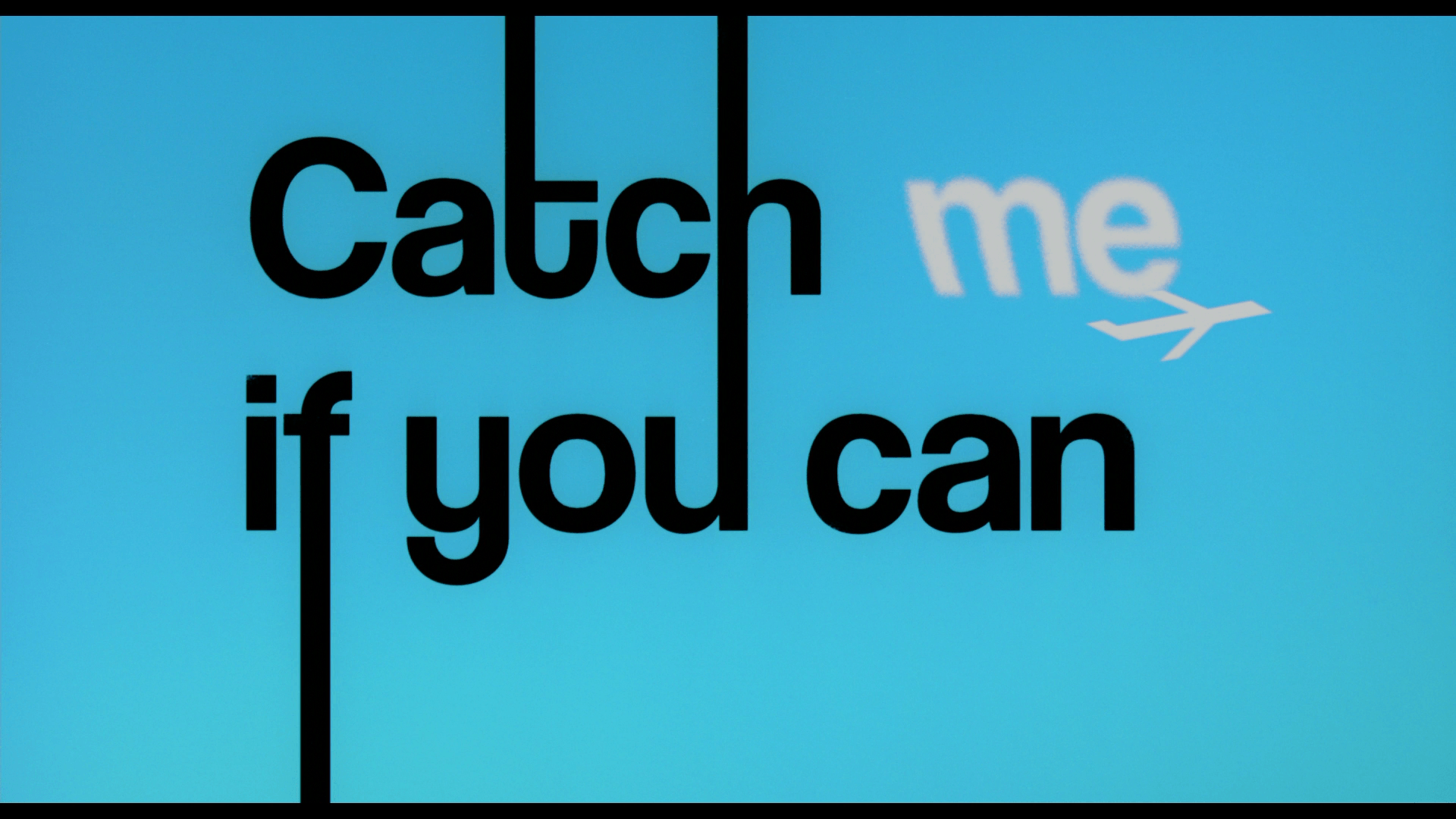
Please (i) report the optimal parameter values, and (ii) compute the Confusion matrix and report the sensitivity, specificity, and the overall accuracy using the testing data.

1. Now we shall build the best classifier to predict whether a check is forged or not using the training data and the SVM with polynomial basis kernel. We shall find the optimal tuning parameters C, degree and scale by using the command line:

tuneLength = 4

Please (i) report the optimal parameter values, and (ii) compute the Confusion matrix and report the sensitivity, specificity, and the overall accuracy using the testing data.

1. Which SVM classifier, namely, (1) linear SVM, (2) SVM with radial basis kernel, and (3) SVM with polynomial kernel, give us the best overall accuracy for the Titanic data?



<https://wallpapercave.com/catch-me-if-you-can-wallpapers>