

# STA 4364 HW 3

**Submission Format:** Please submit your homework as 1) a HTML or pdf document, and 2) also submit the source file in either R Markdown or Jupyter notebook format (at most one of each type of file).

Problems can be done in Python or R. ISL = Introduction to Statistical Learning textbook.

**Problem 1** In this problem, we will examine the German Credit dataset that can be found on Webcourses with the homework in the file `SouthGermanCredit.asc`. All the column names are in German, but you can find the English translations of the columns at this site. We are interested in the `kredit` response, which indicates if an individual has fulfilled their credit contract. Analyze this dataset by following the steps below:

- (a) Load the data using `read.table()`. Rename the columns with their English names. Split the data into a training and test set.
- (b) Perform a logistic regression using the full set of features. Comment on relevant features. Narrow down your features into the most relevant predictors. What are they? Create a reduced model using the set of features you have identified.
- (c) Plot an ROC curve and calculate the AUC of your curve for the full and reduced model on both the training and test set (4 ROC curves in all). Comment on the accuracy and overfitting that you observe for the full and reduced models.

**Problem 2** Analyze the dataset in Problem 1 using LQA and QDA. You should report:

- Summary of each model
- The ROC curve and the AUC of each model
- The comparison among LDA, QDA and logistic regression