Assignment #7

1. LDA
    a. LDA stands for Linear Discriminant Analysis. It is a supervised learning method in which each observation is associated with a class label.
    b. LDA is most useful with high-dimensional data, as it is used to reduce dimensionality.
2. Different from logistic regression?
    a. LDA assumes a multivariate normal distribution, while logistic regression makes no assumptions.
    b. Logistic regression models are for more flexible; LDA implies strict assumptions.
3. ROC
    a. ROC stands for receiving operating characteristic, and it is a method generally used to represent and evaluate nonbinary classification models in data.
4. Specificity and sensitivity
    a. Sensitivity: i.e., the true positive rate; depending on the classifier, sensitivity helps us determine the number of "true positive" cases in the ROC
    b. Specificity: i.e., the false positive rate; 1 – specificity
        i. Measuring the proportion of true/actual negative cases that are incorrectly identified as "positive" by our model.
    c. More important?
        i. It depends on the nature of the data being used as well as the domain. Whether one should be prioritized over the other depends entirely on the specific goal of the analyst.
5. Calculate prediction error rate
    a. (FP+FN) divided by the total number of cases (9,896)