

## Math 5364 Homework 10

1. Split `germancredit.csv` into 70% training and 30% test data. Create models for predicting default using the following learning algorithms, and find the area under the ROC curve for each model.
  - (a) Decision tree.
  - (b) Weighted  $k$ -nearest neighbors.
  - (c) Naive Bayes
2. Consider the following cost structure
  - Predicting someone will default when they don't: \$5,000 in lost interest payments.
  - Predicting someone will not default when they do: \$20,000 in lost principle investment.
  - (a) Given this cost structure, what is the optimal value for the probability threshold  $p_0$ ?
  - (b) Calculate the total cost for each of the models in problem 1 on the test data using the optimal threshold.
  - (c) Calculate the total cost for each of the models in problem 1 on the test data using a threshold of 0.5.