Math 583 HW 11 2023 Due Wednesday, Nov. 29.
Final: Wednesday, Dec. 13, 2:45-4:45. Exam 3, Wed. Dec. 6.

In the Math lab, the computers in the back, 10)–25), tend to have the needed R packages. If you have R on your personal computer, you may need to install packages the first time you use a given computer. Use the `install.packages("e1071")` and `install.packages("randomForest")` commands near the top of the *hdrhw* file, pick a mirror from Iowa or Indiana, and follow the directions.

A) 5.15 This problem creates a classification tree. The dataset has $n = 81$ children who have had corrective spinal surgery. The variables are $Y = Kyphosis$: postoperative deformity is present/absent, and predictors $x_1 = Age$ of child in months, $x_n = Number$ vertebrae involved in the operation, and $Start$ = beginning of the range of vertebrae involved.

a) Use the R code for this part to print the classification tree. Then predict whether $Y = absent$ or $Y = present$ if $Start = 13$ and $Age = 25$.

b) Then predict whether $Y = absent$ or $Y = present$ if $Start = 10$ and $Age = 120$. Note that you go to the left of the tree branch if the label condition is true, and to the right of the tree branch if the label condition is not true.

Problem numbers and example numbers are from the Olive text. Do the two source commands to get the data into R.

B) 5.16 This is the pottery data of Problem 5.11, but the 28 cases were classified as Arrentine for $y = -1$ and nonArrentine for $y = 1$.

a) Copy and paste the commands for this part into R. These commands make the data and do bagging. Copy and paste the truth table into *Word*. What is the AER?

b) Copy and paste the commands for this part into R. These commands do random forests. Copy and paste the truth table into *Word*. What is the AER?

c) Copy and paste the commands for this part into R. These commands do SVM with a fixed cost. Copy and paste the truth table into *Word*. What is the AER?

d) Copy and paste the commands for this part into R. These commands do SVM with a cost chosen by 10-fold CV. Copy and paste the truth table into *Word*. What is the AER?

C) 5.17 This problem uses the Gladstone brain weight data and classifies gender (F for $y = -1$, M for $y = 1$) using various predictors including head measurements, brain weight, and height. Some outliers were removed and the data set was divided into a training set with $n = 200$ cases and a test set with $m = 61$ cases.

a) Copy and paste the commands for this part into R. These commands make the data and do bagging. Copy and paste the truth table into *Word*. What is the AER?

b) Copy and paste the commands for this part into R. These use bagging on the training data and validation set. Copy and paste the truth table into *Word*. What is the bagging validation error rate?

c) Copy and paste the commands for this part into R. These commands do random forests. Copy and paste the truth table into *Word*. What is the AER?

d) Copy and paste the commands for this part into R. These use random forests on the training data and validation set. Copy and paste the truth table into *Word*. What

is the random forests validation error rate?

e) Copy and paste the commands for this part into *R*. These commands do SVM with a cost chosen by 10-fold CV. Copy and paste the truth table into *Word*. What is the AER?

f) Copy and paste the commands for this part into *R*. These commands do SVM with a cost chosen by 10-fold CV on the training data and validation set. Copy and paste the truth table into *Word*. What is the SVM validation error rate?