# Quiz 7. AMS 597

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

# The quiz is due at the end of the lecture by 11:20am – please submit no later than 11:20am. Please email your completed quiz to your TA at: [song.jiecheng@stonybrook.edu](mailto:song.jiecheng@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 emailed your solutions.

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

#### CART with the Caesarian Data – Classification Task

Ever thought about the future with ***Robot Doctors***? Now imagine, a pregnant woman would visit such a robot doctor, giving it answers to a few questions, and then the robot doctor will tell her whether she should have a Caesarian delivery or not. Now think again, how about you build such a robot doctor by yourself, right now!

The caesarian.csv dataset attached contains 6 variables and 80 observations:

* *Age*: Age of each pregnant woman
* *Delivery number:* Values{1, 2, 3, 4} indicating which delivery it is for each woman
* *Delivery time*: Values {0, 1, 2} indicating {0 = timely, 1 = premature, 2 = latecomer}
* *Blood Pressure*: Values {2, 1, 0} indicating {0 = low, 1 = normal, 2 = high}
* *Heart Problem*: Values {1, 0} indicating {0 = apt, 1 = inept}
* *Caesarian*: Values {0, 1} indicating {0 = No, 1 = Yes}

Your task is to split the data randomly into training (75%) and testing (25%), first build the full tree and then establish an optimal model with pruning and 10-fold cross validation using the training data, and then use that model to predict whether each woman in the testing data should have a Caesarian delivery or not. Please use *rpart* to build the classification tree and *rattle* to plot the tree.

1. Please use the random seed 123 to divide the dataset into 75% training and 25% testing.

1. Please first build a fully grown tree using the training data, and draw the tree plot using *rattle*. Next please use this tree to predict the delivery mode (Caesarian or not) for each woman in the testing data. Please compute the Confusion matrix and report the sensitivity, specificity and the overall accuracy for the testing data.
2. To make the tree more robust, we will prune the fully grown tree using the training data with 10-fold cross-validation. Please (1) show the complexity plot, (2) report the best CP value, and (3) draw the pruned tree using *rattle*.
3. Please use this optimal pruned tree to predict the delivery mode (Caesarian or not) for each woman in the testing data. Please compute the Confusion matrix and report the sensitivity, specificity and the overall accuracy for the testing data.