**DSBA/MBAD 6211 Assignment 2**

Due: 11:59pm @ 10/14/2020

We will analyze the Boston housing dataset to classify and predict ***whether a property exceeds the median value of owner-occupied home in Boston***. Please only use the binary DV, and discard the continuous DV. Please apply both Random Forest and Support Vector Machines to analyze the dataset.

* **Variable and model naming requirements:**
  + Please include your ***name initials*** to the data frame names as well as model names in your R coding.
  + Please instance, in my coding, I would name the data frames as ***dfKZ, dfKZ.train***, and ***dfKZ.valid.*** I would also name the models as ***RFKZ, SVMKZ***, etc.

Please submit a Word document including answers to the following questions.

1. Summarize the random forest model results

* What is the optimal parameter mtry?
* Please provide the variable importance ranking. Do they make sense to you?
* Please provide the confusion matrix for the test dataset

1. For SVM, please compare results for two kernel functions, linear vs. radial. Which kernel function does lead to a better model performance in the default setting?
2. Please tune the SVM model with selected kernel function, and identify best parameters.

* Please provide the confusion matrix for the test dataset
* What is the optimal parameter(s) for the selected kernel function

1. Overall, which method (random forest vs. SVM) is performing better?
2. Copy and paste your R codes at the end of the WORD document.

Data dictionary:

* CRIM per capita crime rate by town
* ZN proportion of residential land zoned for lots over 25,000 sq.ft.
* INDUS proportion of non-retail business acres per town.
* CHAS Charles River dummy variable (1 if tract bounds river; 0 otherwise)
* NOX nitric oxides concentration (parts per 10 million)
* RM average number of rooms per dwelling
* AGE proportion of owner-occupied units built prior to 1940
* DIS weighted distances to five Boston employment centers
* RAD index of accessibility to radial highways
* TAX full-value property-tax rate per $10,000
* PTRATIO pupil-teacher ratio by town
* LSTAT % lower status of the population
* MEDV Median value of owner-occupied homes in $1000
* CAT.MEDV binary indicator whether a home’s value exceeds the median, 1 yes and 0 no