DATS 7860 Statistical and Machine Learning for Big Data

Homework 6 (Chapter 8)

1. The random forest method is implemented in the R packages randomForest and randomForestSRC. You have tried the randomForest package in the lab exercise. Now please read the introduction webpage of the randomForestSRC package (uploaded to D2L as "randomForestSRC_Introduction.pdf", which is also available at: https://www.randomforestsrc.org/articles/getstarted.html). Write a few sentences summarizing what additional functionalities randomForestSRC has implemented compared with randomForest. Particularly, please answer what types of data / models can randomForestSRC handle in addition to the regression and classification problems we discussed in class.

Problems below are all from Exercises 8.4 of Chapter 8 of the textbook (ISL):

2. Problem 4 (Page 362).

3. Problem 8 (Page 363).

Some specific notes for this problem:

- Part (e): To answer this question - "describe the effect of m, the number of variables considered at each split, on the error rate obtained", please try at least 3 different values of *m* and check their corresponding test MSEs.

- Please submit your complete R codes for this problem.

4. Problem 9 (Page 363).

Some specific notes for this problem:

- Please submit your complete R codes for this problem.

5. Problem 11 (Page 364).

Some specific notes for this problem:

- Part (c): for this question “How does this compare with the results obtained from applying KNN or logistic regression to this data set”, you can either apply the KNN or logistic regression, or both of them.

- Please submit your complete R codes for this problem.