BITS F464 - Machine Learning
II Semester 2018-2019
Assignment #1
Weightage: 10%
Due Data: 06th April, 2019

Heart Disease Prediction

A set of 14 of attributes were selected, out a total of 76 attributes, to predict heart disease in certain population (Cleveland, US). The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4 (indicating extreme susceptibility).

Use different Machine Learning techniques and algorithms to predict the class label "goal".
The dataset has been uploaded on NALANDA in csv format. A brief description of the dataset is available on Kaggle and Data World.

Part - I

Apply the following classification techniques to solve the above problem:

1. NBC
2. Logistic Regression
3. SVM – Linear hard/soft margin, Non-linear SVM
4. Decision Trees
5. Discriminant Functions – Least Square, FLD, Perceptron
6. Ensembles - Random forest and Adaboost

Compare the accuracies of different classifiers.

Part – II

1. Change the data labels to just reflect the presence/absence of heart disease.
2. Solve the resultant binary classification problem.
3. How can you to retrieve the original class labels (2,3,4) from the class label – 1?
    a. Compare the accuracy of your approach using the original data as ground truth.
    b. Report your findings

Implement all your models in R.

Assignment Groups: Form a team of up to 3 students.
You will be provided with a presentation template in which you will need to fill the details.

Navneet Goyal/Kamlesh Tiwari
goel@/kamlesh.tiwari@