C-M-005: Machine Learning Lab I '21-22

Lab. 4 (Given November 20, 2021; Due November 22, 2021)

Your answers must be entered in LMS by midnight of the day it is due. You must submit the code as well as a self-contained PDF which has the approach, an explanation of the implementation, the output as well as anything else asked by the question. Marks devoted to this Lab. are indicated in the "Syllabus" sheet that was provided the first day of class.

---

1. Consider the Wisconsin Breast Cancer dataset available from

   `http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)`

   The dataset has 32 attributes that predict malignancy. There are a total of 569 data patterns. Use 5-fold cross-validation. (i) Use Keras or any other framework to construct a decision tree from the training data and obtain the performance on the test data, (ii) Construct a random forest (of say, 100 trees) from the training data and use the random forest to obtain the performance on the test data, (iii) Compare the performance you obtain in (i) and (ii) **(20+20+10 points)**

2. (Optional) Repeat the exercise but add $\pm 10\%$ noise to 25% of the data