**Instructions Lab 13**

In this lab you will experiment with Random Forest. In earlier lab you have worked on the Breast Cancer dataset. You can use the same dataset today.

1. Download the data from

<https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/>

2. you can select the number of features that you want in each tree. Start with two features and build a random forest. Check how the accuracy increases as the number of trees is increased in the forest. Note the best accuracy obtained.

3. Repeat step 2 for increasing number of features.

4. Plot a graph of best accuracy (for a given number of features) versus number of features.

5. Use the above graph to decide the optimal number of features for this dataset.

6. Using the optimal number of features find the importance of each feature using OOB samples.

7. Find the MI values of all the features.

8. Compare the results obtained in 7 and 8

Are the two lists identical? Can you find if they are strongly correlated?

*Show your work, even if it is partial, since each lab is graded based on what you achieve during the lab.*