Consider the two data files attached to this assignment. This data is taken from the UCI Machine Learning Repository and contains biomechanical features of some patients (http://archive.ics.uci.edu/ml/datasets/vertebral+column). The task is to predict whether the patient is normal or abnormal (call it Data1). The second dataset splits the abnormal group into two different diagnoses (call it Data2). Both datasets have same 310 feature vectors, six features, and a class column. Perform the following tasks with these datasets and submit the answers asked for in each task listed below.

Use ***fitctree*** function of Matlab to solve the problem.

Each submission must be individual work of each student. Any plagiarism detected will be severely punished.

1. Take Data1 and split it into randomly selected 210 training instances and remaining 100 as test instance. Create decision trees using the training set and the "minimum records per leaf node" values of 5, 10, 15, 20, and 25. [**30**]
    a. Show the tree for the value 25. Comment on what you notice about the five trees.
    b. For each tree compute and report the accuracy, precision, and recall (***We will be studying these terms in tomorrow's lecture***) values. Comment on the comparison of these values and show these values on a plot.
    c. Now limit yourself to the case of 10 minimum records per leaf node. Repeat the tree learning exercise five times by randomly choosing different sets of 210 training instances. Report the accuracy, precision, and recall values for each run and also their averages and standard deviations. Comment on the variability of the values as the random sample changes.
2. Repeat the same tasks as done in Question-1 above for Data2. In addition to reporting results for 2a, 2b, and 2c, comment on the comparison of results obtained for 1c and 2c. Give your analysis for the differences in results. Label this answer as 2d. [**30**]
3. Take Data1 for this question. Partition each column into four sets of equal widths of values. Assign these intervals as values 0, 1, 2, and 3 and replace each value by its corresponding interval value. [**30**]
    a. Show the boundaries for each interval for each attribute.
    b. Learn a decision tree with this transformed data and compute performance parameters in the same way as done for 1c and 2c.
    c. Compare these results with those obtained for 1c. Analyze the differences in performance and give your intuitive reasons why these differences are observed.
4. Extra **10** points are for good organization and presentation of results in your submission. Submit all the results in a single file.