MACHINE INTELLIGENCE AND EXPERT SYSTEMS

AUTUMN SEMESTER – 2019

Computer Assignment on Decision Trees

Q1. TRAIN AND TEST A DECISION TREE CLASIFIER WITH AND WITHOUT PRUNING

- You are supposed to train a decision tree to help someone in prediction of the class of Iris plant.
- The file named "iris_train_data.csv" has the training data and file named "iris_test_data.csv" has the test data.
- Read data from csv files using "PANDAS" library of python
- Use the training data to train the Decision Tree Classifier from "SKLEARN" library of python and calculate the training accuracy of the classifier.
- Use "entropy" as "criterion" to choose Information Gain for measurement of quality of split, in DecisionTreeClassifier Configuration
- Print the depth and number of leaf nodes in learnt tree
- Calculate the test accuracy with the help of test dataset.
- Use pruning during training to avoid over-fitting by changing the Decision Tree Classifier parameters like
    - max_leaf_nodes : Reduce the number of leaf nodes,
    - max_depth : Reduce the depth of the tree to build a generalized tree.

  And calculate the test accuracy in above 2 cases of pruning.

NOTE: For accuracy calculation, write your own function. Use of any imported module for accuracy calculation is not allowed.

- **Details about the Dataset:**
    1. Title: IRIS Dataset
    2. Relevant Details:
        a. The dataset provides 3 classes of IRIS plant.
        b. Predicted Attribute: Class of Iris plant
        c. Number of Attributes: 4 (Al Real Valued)
        d. Attribute Information:
            i. sepal_length -> sepal length in cm
            ii. sepal_width -> sepal width in cm
            iii. petal_length -> petal length in cm
            iv. petal_width -> petal width in cm
        e. Class Information:
            i. Iris Setosa
            ii. Iris Versicolour
            iii. Iris Virginica