In the Name of God
Machine Learning (Spring 2013)
Instructor: Dr. Sattar Hashemi
Homework #3: MDL discretization & decision tree
Due Date: April 9th

Discretization is a process of converting or partitioning continuous attributes, features or variables to discretized or nominal attributes/ features/ variables/intervals.

You are familiar with MDL discretization algorithm which was proposed by Fayyad & Irani. In this homework we want you to implement the MDL discretization algorithm on some continuous datasets and implement a decision tree which is given in the enclosed paper.

You can use at least 4 toy datasets of UCI (including Wine, Iris, Glass, and Heart datasets or other suitable datasets) in this homework. For the decision tree, you can use the given pseudo code of *ID3* in your text book (Tom Michael's book).

Your code shouldn't handle missing values or use pruning methods. It has to contain just the Fayyad & Irani MDL discretization method and typical ID3 of your book.

Please write a report for your homework in one page and compare your results in each dataset. You should calculate train and test accuracies in each dataset. Also you should calculate Precision, Recall, and F-measure for minority class, the class with the lowest number of instances. For example consider a dataset that contains 3 classes of instances and the following distribution.

Class	# of instances
C1	48
C2	59
C3	71

In this dataset class C1 with 48 instances is the minority class. Use 10 times five-fold cross validation to separate data into train and test sets. Report all train and test measures in a table over all datasets.

➢ Please put your files and codes in a folder and upload the **.zip** file on "HWS", considering the due date.
➢ Please put your files in a folder and upload the **.zip** file on "hws", considering the due time. **The name of the uploading file should be your lastname_firstname.**