

**QUESTION 2 (5pts)**

- 1. Define classification data**
  
- 2. What is a classifier**
  
- 3. Define 4 stages of the process of building a Classifier**

## Part - 1    Question - 2

### ① Classification Data :

Classification Data can be defined as a data set, which contains data tuples (or) records from given data along with their associated class labels.

Classification Data format:

- The Data table of classification Data has its key attribute removed.
- A special attribute, called as class attribute must be distinguished.
- The values of the class attribute are called as class labels.
- Class attributes are categorical, each value serves as a category, or a class.
- The class labels are discrete and unordered.
- A data tuple (or) record in classification data has attribute part and class part.
- The attribute part is called data tuple, or

attribute vector, data vector, sample, example, instance, data point (with associated class label).

## ② Classifier:

A classifier is a black-box that is used to classify records for which the class label is unknown.

A classifier is a final product of a learning process that uses a classification dataset and a classification algorithm.

- A classifier is the end result of a process that uses training data and testing data and a classification algorithm to generate patterns that can classify new data.

The need for refinement, effectiveness and completion of a classifier is determined based on its predictive accuracy.

- We terminate the process of building a classifier if it has been trained and tested and the predictive accuracy is on acceptable level.
- The patterns that a classifier uses take different forms,

such as ensembles of trees, trained networks like neural or Bayesian.

### ③ Building a Classifier:

Two data sets, namely training data set and testing data set have to be created in order to build a classifier. These two data sets must be disjoint sets.

Training and Test data may differ in nature but must have same format.

Stages of building a classifier:

Stage 1 - Training:

In this stage, a classification algorithm uses the training data to build a basic structure for set of patterns.

These patterns can be called as discriminant and/or characteristic rules or other pattern descriptors. This structure can be called as basic or learned classifier (or) model.

Using these patterns we can classify unknown sets of objects (unknown tuples/records).

### Stage 2 - Parameter tuning:

After building a basic or learned classifier, the parameter settings need to be optimized.

By testing the classification model using a dataset which is already used for training called as validation dataset, an error rate can be seen.

For example: Resubstitution (N;N) method.

This error rate expresses how good or bad the results are on the training data. It reflects the imprecision of training results. Lower the error rate better will be the classifier.

### Stage 3 - Testing:

In this stage, test data is used to evaluate the classifier which was built by predicting the class values of it.

## Predictive Accuracy:

Predictive Accuracy of the model is the percentage of well classified data in testing data set.

Predictive accuracy gets computed in this stage.

If the predictive accuracy is higher the the model is good.

## Stage 4 - Consolidation :

In this stage, consolidation of stages 1, 2 and 3 occurs in order to build a classifier as the final product.