MES College of Engineering Pune-01
Department of Computer Engineering

Name of Student:	**Class:**
Semester/Year:	**Roll No:**
Date of Performance:	**Date of Submission:**
Examined By:	**Subject: LPVI (E-VI BI)**

Assignment No. 5

Aim: Perform the data classification using classification algorithm. Or perform the data clustering using clustering algorithm

Theory:

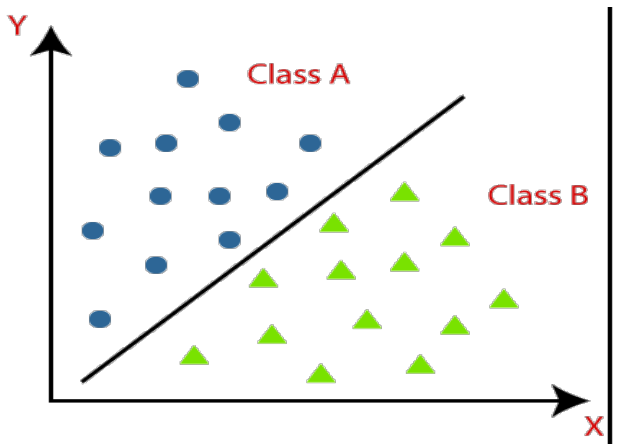
Data Classification?

Classification algorithms are used to categorize data into a class or category. It can be performed on both structured or unstructured data. Classification can be of three types: binary classification, multiclass classification, multilabel classification.

What is the Classification Algorithm?

The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, **Yes or No, 0 or 1, Spam or Not Spam, cat or dog,** etc. Classes can be called as targets/labels or categories.

Classification algorithms can be better understood using the below diagram. In the below diagram, there are two classes, class A and Class B. These classes have features that are similar to each other and dissimilar to other classes.



The algorithm which implements the classification on a dataset is known as a classifier. There are two types of Classifications:

- **Binary Classifier:** If the classification problem has only two possible outcomes, then it is called as Binary Classifier.

  Examples: YES or NO, MALE or FEMALE, SPAM or NOT SPAM, CAT or DOG, etc.

- **Multi-class Classifier:** If a classification problem has more than two outcomes, then it is called as Multi-class Classifier.

  Example: Classifications of types of crops, Classification of types of music.

What are clustering algorithms?

Clustering is an unsupervised machine learning task. You might also hear this referred to as cluster analysis because of the way this method works. Using a clustering algorithm means you're going to give the algorithm a lot of input data with no labels and let it find any groupings in the data it can. Those groupings are called clusters. A cluster is a group of data points that are similar to each other based on their relation to surrounding data points. Clustering is used for things like feature engineering or pattern discovery.

Types of clustering algorithms

There are different types of clustering algorithms that handle all kinds of unique data.

Density-based

In density-based clustering, data is grouped by areas of high concentrations of data points surrounded by areas of low concentrations of data points. Basically, the algorithm finds the places that are dense with data points and calls those clusters.

Distribution-based

With a distribution-based clustering approach, all of the data points are considered parts of a cluster based on the probability that they belong to a given cluster. It works like this: there is a center-point, and as the distance of a data point from the center increases, the probability of it being a part of that cluster decreases.

Centroid-based

Centroid-based clustering is the one you probably hear about the most. It's a little sensitive to the initial parameters you give it, but it's fast and efficient. These types of algorithms separate data points based on multiple centroids in the data. Each data point is assigned to a cluster based on its squared distance from the centroid. This is the most commonly used type of clustering.

Hierarchical-based

Hierarchical-based clustering is typically used on hierarchical data, like you would get from a company database or taxonomies. It builds a tree of clusters so everything is organized from the top-down.This is more restrictive than the other clustering types, but it's perfect for specific kinds of data sets.

Questions **https://www.javatpoint.com/classification-algorithm-in-machine-learning**

1. **What do you mean by Clustering?**
2. **What is a Hierarchical Clustering Algorithm?**
3. **Explain the Differences between Classification and Clustering?**
4. **What Is classification?**
5. **How to Evaluated a classification model?**