

Unit– V Unsupervised Machine Learning Models

5.1.1 Introduction of Unsupervised Learning

❖ Brief explanation of unsupervised Machine Learning

What is Unsupervised Learning?

- Unsupervised learning is a machine learning technique in which models are not supervised using training dataset.
- Instead, models itself find the hidden patterns and insights from the given data.
- It can be compared to learning which takes place in the human brain while learning new things.
- **Definition:** Unsupervised learning is a type of machine learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision.

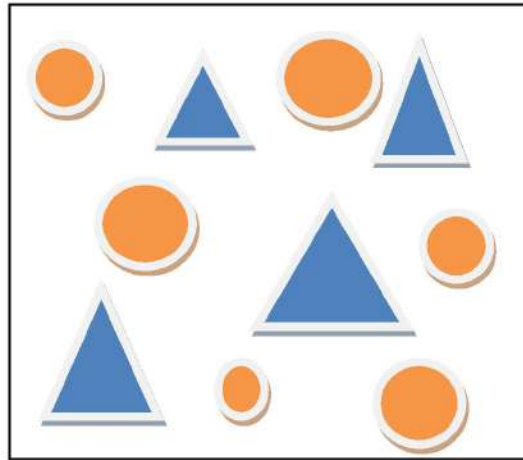
❖ Need of unsupervised learning

- Unsupervised learning (also known as knowledge discovery) uses unlabeled, unclassified, and categorized training data.
- The main goal of unsupervised learning is to discover hidden and interesting patterns in unlabeled data.
- It is helpful for finding useful insights from the data.
- It is much similar as a human learns to think by their own experiences, which makes it closer to the real AI.
- It works on unlabeled and uncategorized data which make unsupervised learning more important.
- In real-world, we do not always have input data with the corresponding output so to solve such cases, we need unsupervised learning.

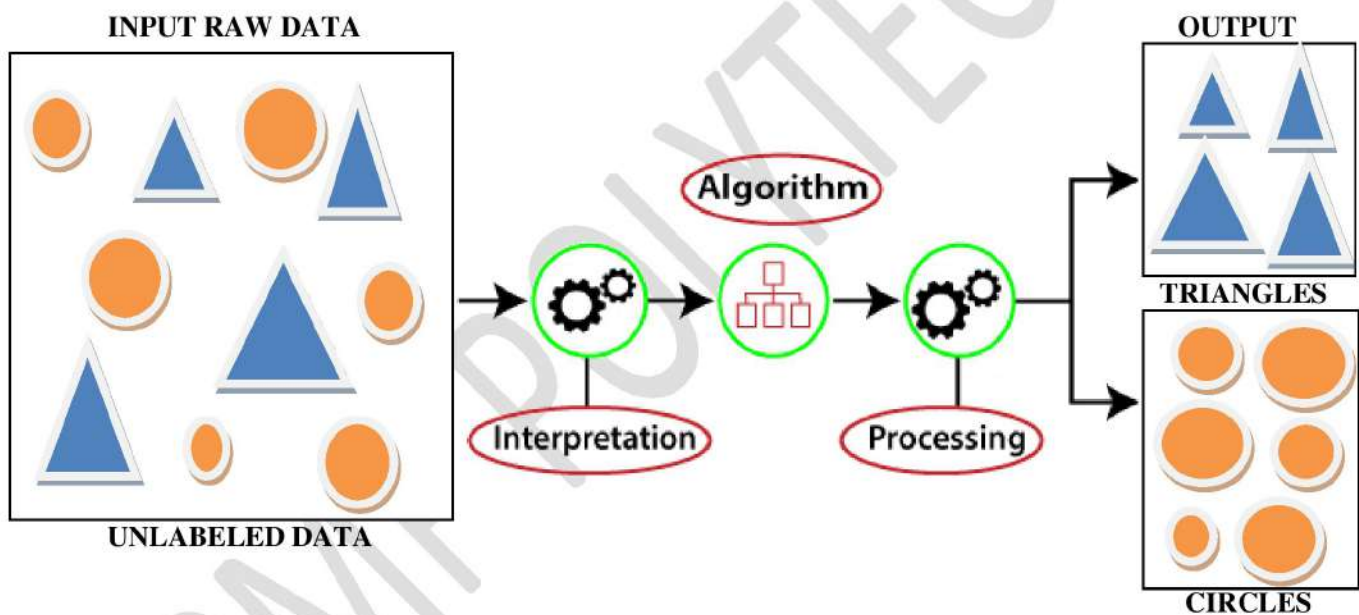
❖ Real world examples of unsupervised Learning

Example: Suppose the unsupervised learning algorithm is given an input dataset containing images of different types of triangles and circles. The algorithm is never trained upon the given dataset, which means it does not have any idea about the features of the dataset.

The task of the unsupervised learning algorithm is to identify the image features on their own. Unsupervised learning algorithm will perform this task by clustering the image dataset into the groups according to similarities between images.



Working of Unsupervised Learning



- Here, we have taken an unlabeled input data, which means it is not categorized and corresponding outputs are also not given.
- Now, this unlabeled input data is fed to the machine learning model in order to train it.
- First, it will interpret the raw data to find the hidden patterns from the data and then will apply suitable algorithms such as k-means clustering, Decision tree, etc.
- Once it applies the algorithm divides the data objects into groups according to the similarities and difference between the objects.

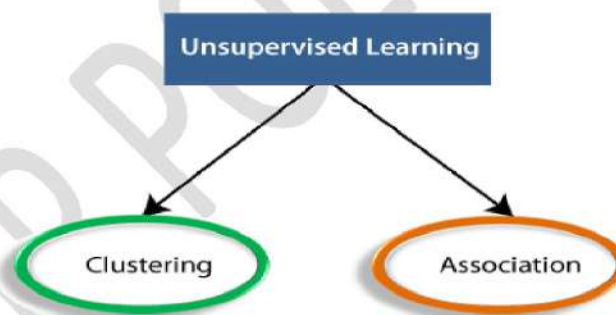
- **List unsupervised learning algorithms**

Below is the list of some popular unsupervised learning algorithms:

- **K-means clustering**
- **KNN (k-nearest neighbors)**
- **Hierarchical clustering**
- **Anomaly detection**
- **Neural Networks**
- **Principle Component Analysis**
- **Independent Component Analysis**
- **Apriori algorithm**
- **Singular value decomposition**

5.2.1 Types of Unsupervised Learning

- ❖ **Clustering: Definition, list clustering methods, list real world applications/examples (fruits, vegetables, computer devices (input and output etc))**



- **Clustering:** Clustering is a method of grouping the objects into clusters such that objects with most similarities remain into a group and have less or no similarities with the objects of another group.
- Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.
- 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.



sample



Cluster/group

- **Example:** Let's understand the clustering technique with the real-world example of Mall: When we visit any shopping mall, we can observe that the things with similar usage are grouped together. Such as the t-shirts are grouped in one section, and trousers are at other sections, similarly, at vegetable sections, apples, bananas, Mangoes, etc., are grouped in separate sections, so that we can easily find out the things. The clustering technique also works in the same way. Other examples of clustering are grouping documents according to the topic.
- The clustering technique can be widely used in various tasks. Some most common uses of this technique are:
 - Market Segmentation
 - Statistical data analysis
 - Social network analysis
 - Image segmentation
 - Anomaly detection, etc.

❖ Association: Definition, list association methods, list real world applications/examples

- **Association:** An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database.
- It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective.
- Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item.
- A typical example of Association rule is Market Basket Analysis.
- The association rule learning is one of the very important concepts of machine learning, and it is employed in **Market Basket analysis, Web usage mining, continuous production, etc.**

- **Example:** Here market basket analysis is a technique used by the various big retailer to discover the associations between items. We can understand it by taking an example of a supermarket, as in a supermarket, all products that are purchased together are put together.
- For example, if a customer buys bread, he most likely can also buy butter, eggs, or milk, so these products are stored within a shelf or mostly nearby. Consider the below diagram:



❖ Advantage and Disadvantage of unsupervised learning algorithm

Advantages of Unsupervised Learning

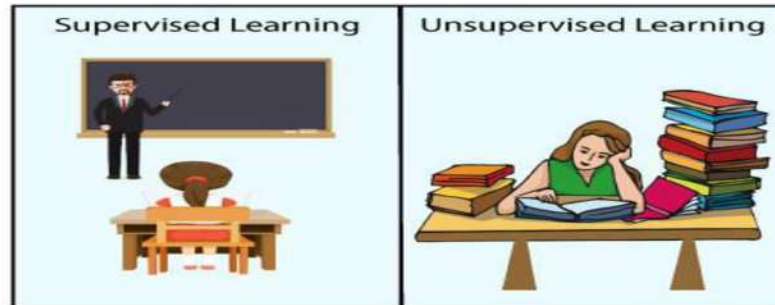
- Unsupervised learning is used for more complex tasks as compared to supervised learning because, in unsupervised learning, we don't have labeled input data.
- Unsupervised learning is preferable as it is easy to get unlabeled data in comparison to labeled data.

Disadvantages of Unsupervised Learning

- Unsupervised learning is intrinsically more difficult than supervised learning as it does not have corresponding output.
- The result of the unsupervised learning algorithm might be less accurate as input data is not labeled, and algorithms do not know the exact output in advance.

5.3.1 Differentiate Supervised and Unsupervised Learning

- Supervised and Unsupervised learning are the two techniques of machine learning.
- But both the techniques are used in different scenarios and with different datasets.
- Below the explanation of both learning methods along with their difference table is given.



Supervised Learning	Unsupervised Learning
Algorithms are trained using labeled data.	Algorithms are trained using unlabeled data.
Model takes direct feedback to check if it is predicting correct output or not.	Model does not take any feedback.
Model predicts the output.	Model finds the hidden patterns in data.
Input data is provided to the model along with the output.	Only input data is provided to the model.
Goal: To train the model so that it can predict the output when it is given new data.	Goal: To find the hidden patterns and useful insights from the unknown dataset.
It needs supervision to train the model.	It does not need any supervision to train the model.
Categorized: Classification and Regression	Categorized: Clustering and Association
Here, we know the input as well as corresponding outputs.	Here, we have only input data and no corresponding output data.
Supervised learning model produces an accurate result.	Unsupervised learning model may give less accurate result as compared to supervised learning.
IT is not close to true Artificial intelligence.	It is more close to the true Artificial Intelligence.
It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc.	It includes various algorithms such as Clustering, KNN, and Apriori algorithm.