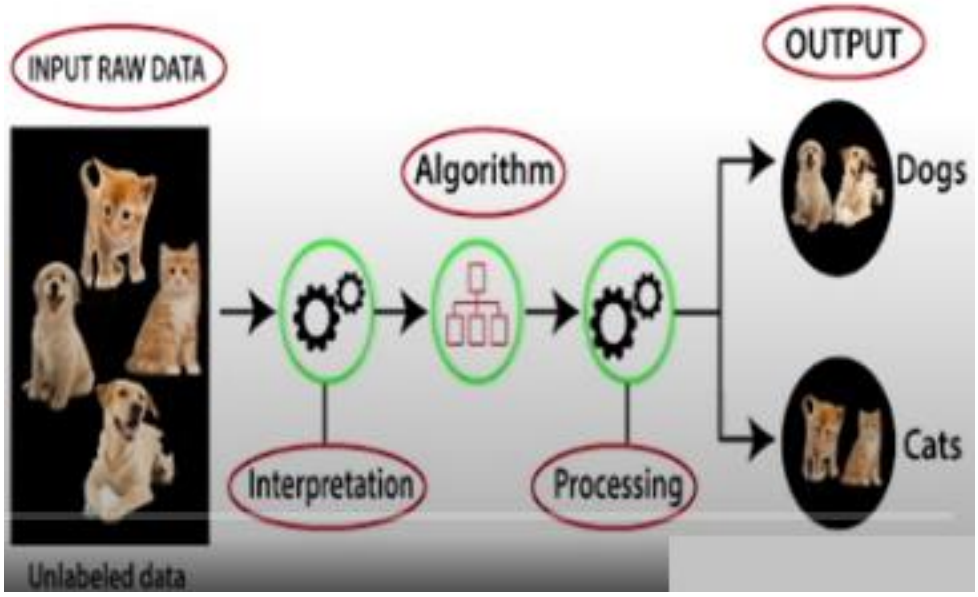
UNIT-V

ML Unsupervised Learning

Unit V: ML Unsupervised Learning	5a. Explain an unsupervised algorithm. 5b.Compare clustering and association unsupervised algorithms. 5c. List key points of K-means clustering algorithm. 5d.List and explain the given applications of AI and ML in robotics.	5.1Working of unsupervised learning algorithms 5.2 Types: Clustering and Association 5.3 Unsupervised learning algorithms: K-means clustering (key points) 5.4 Application of AI and ML in Robotics-Computer vision, AI enabled manipulation and grasping, AI enhanced navigation and motion control, Real world perception and natural language processing.

Unsupervised Machine learning:

- Learning when there is no hint at all about correct outputs.
- Process of learning in which machine is trained on unlabeled data/uncategorized data without any guidance.
- Find hidden patterns, trends and insights from given data like human brain learn new things by own experiences,which makes it closer to real AI.
- In real world, we donot always have input data with the corresponding output so to solve such cases,we need unsupervised machine learning.



Types of Unsupervised Learning

- **Clustering**
  Clustering is an unsupervised task which involves grouping the similar data points.

- **Association**

  Association is an unsupervised task that is used to find important relationship between data points.

Clustering;

In this technique, which group the unlabeled dataset. A way of grouping the data points into different clusters , consisting of similar data points.

Clustering is the process of dividing the entire data into groups (also known as clusters) based on the patterns in the data.

The object with possible similarities remain in a group that has less or no similarities with another group. It is used by Amazon, Netflix in its recommendations as per the past search of product ,movies respectively.

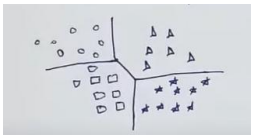
Types of clustering methods

- Partitioning clustering
- Density based
- Distribution model based
- Hierarchical clustering
- Fuzzy clustering

Partitioning clustering: It divides the data into non-hierarchical group . Its I s also known as centroid method .

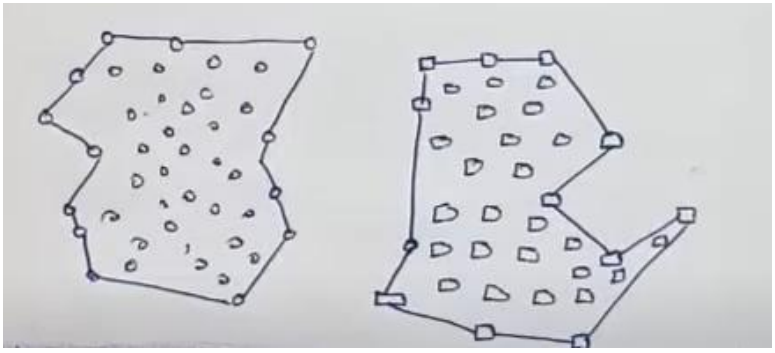
- Data set (Cluster) into set of grroups.
- K- value k=3
- Centroid - based method

Eg . k-means clustering algorithms.



Density based:

It connects highly dense area into clusters and arbitrarily shaped distribution are formed as long as dense region can be connected.



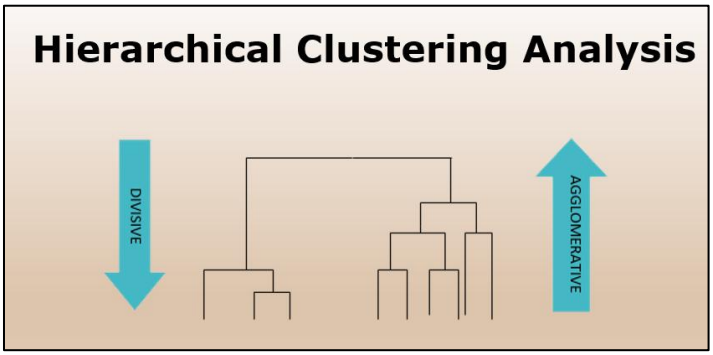
Distribution Model Based:

The data is divided based on probability of how a data set belongs to particular distribution. The grouping is done by assuming some distribution commonly Gaussian Distributions.

Hierarchical Clustering: It Can be used as alternative for partitioned clustering as there is no requirement of pre-specifying no.of clustering to be created. Dataset is divided into clusters to create a **dendrogram** (Tree like).

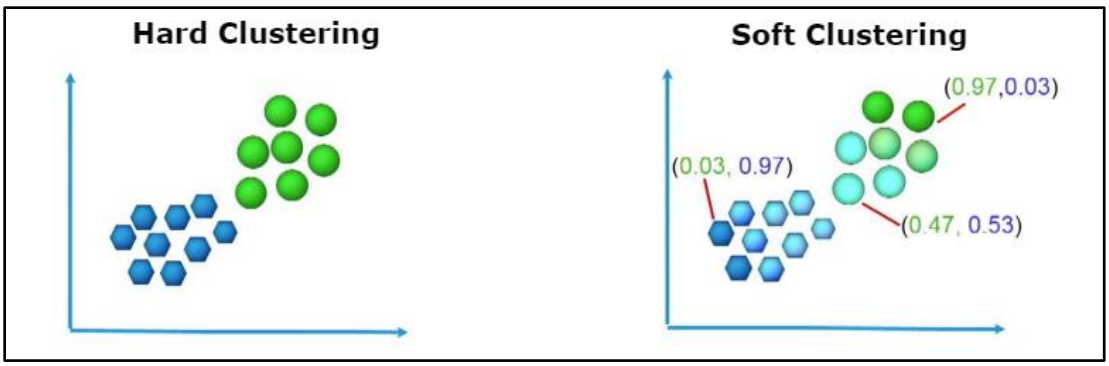
Hierarchical Clustering:

- Set of nesting Clustering
- Organized by Representation tree
- Dendrogram



Fuzzy clustering:

It is type of Soft Method in which data object may belong to more than one cluster. .Each dataset has a set of membership coefficients, which depend on degree of membership to be in a cluster



Exclusive clustering:

- Assign each object to a single group
- Non-overlap clustering

Non exclusive clustering:

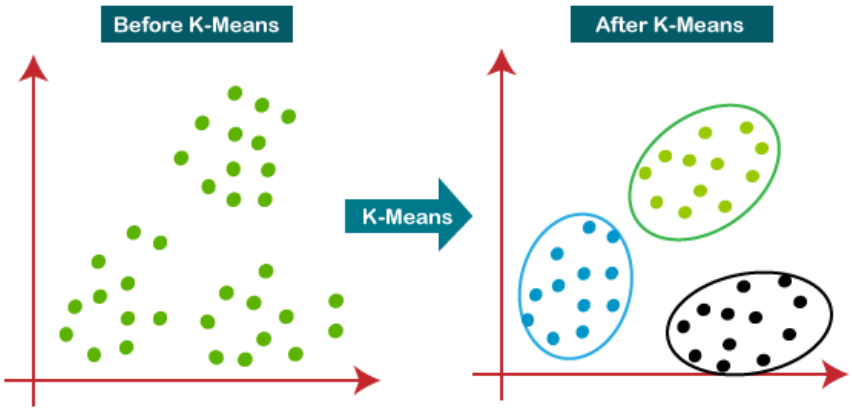
- Object can simultaneously belong one or more than one group
- Overlapping clustering.

Types of Clustering Algorithms

- K-Means clustering.
- Mini batch K-Means clustering algorithm.
- Mean Shift.
- Divisive Hierarchical Clustering.
- Hierarchical Agglomerative clustering.
- Gaussian Mixture Model.
- DBSCAN.
- OPTICS.

K-means clustering:

- K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.
- It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.
- The k-means clustering algorithm mainly performs two tasks:
    - Determines the best value for K center points or centroids by an iterative process.
    - Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.



The working of the K-Means algorithm is explained in the below steps:

Step-1: Select the number K to decide the number of clusters.

Step-2: Select random K points or centroids. (It can be other from the input dataset).

Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.

Step-4: Calculate the variance and place a new centroid of each cluster.

Step-5: Repeat the third steps, which means reassign each data point to the new closest centroid of each cluster.

Step-6: If any reassignment occurs, then go to step-4 else go to FINISH.

Step-7: The model is ready.

Association Rule learning:

It is a type of unsupervised learning technique that checks for dependency of one data item an another data item and maps accordingly so that it can be more profitable.

Types of Association rule learning algorithms:

- Apriori
- Eclat
- F-P GROWTH ALGORITHM

How does Association Rule Learning work?

Association rule learning works on the concept of If and Else Statement, such as if A then B.

Here the If element is called **antecedent**, and then statement is called as **Consequent**. These types of relationships where we can find out some association or relation between two items is known *as single cardinality*. It is all about creating rules, and if the number of items increases,

then cardinality also increases accordingly. So, to measure the associations between thousands of data items, there are several metrics. **These metrics are given below:**

- **Support**
- **Confidence**
- **Lift**

Support

Support is the frequency of A or how frequently an item appears in the dataset. It is defined as the fraction of the transaction T that contains the itemset X. If there are X datasets, then for transactions T, it can be written as:

$$\text{Supp(X)}=\frac{Freq(X)}{T}$$

Confidence

Confidence indicates how often the rule has been found to be true. Or how often the items X and Y occur together in the dataset when the occurrence of X is already given. It is the ratio of the transaction that contains X and Y to the number of records that contain X.

$$\text{Confidence}=\frac{Freq(X,Y)}{Freq(X)}$$

Lift

It is the strength of any rule, which can be defined as below formula:

$$\text{Lift}=\frac{Supp(X,Y)}{Supp(X)\times Supp(Y)}$$

It is the ratio of the observed support measure and expected support if X and Y are independent of each other. It has three possible values:

- If **Lift= 1**: The probability of occurrence of antecedent and consequent is independent of each other.
- **Lift>1**: It determines the degree to which the two itemsets are dependent to each other.
- **Lift<1**: It tells us that one item is a substitute for other items, which means one item has a negative effect on another.

Applications of Association Rule Learning

It has various applications in machine learning and data mining. Below are some popular applications of association rule learning:

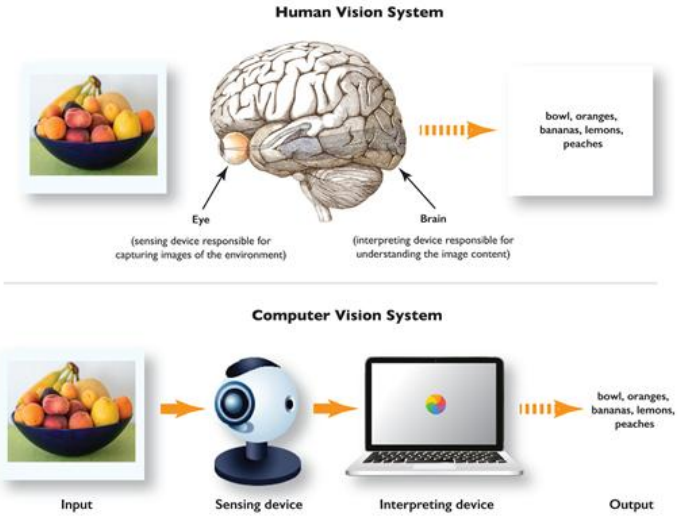
- **Market Basket Analysis:** It is one of the popular examples and applications of association rule mining. This technique is commonly used by big retailers to determine the association between items.
- **Medical Diagnosis:** With the help of association rules, patients can be cured easily, as it helps in identifying the probability of illness for a particular disease.
- **Protein Sequence:** The association rules help in determining the synthesis of artificial Proteins.
- It is also used for the **Catalog Design** and **Loss-leader Analysis** and many more other applications.

Applications of Artificial Intelligence and Machine Learning in Robotics:

1. Computer vision
2. AI enabled manipulation and grasping
3. AI enhanced navigation and motion control
4. Real world perception
5. natural language processing.

1. Computer vision:

Computer vision leverages artificial intelligence (AI) to allow computers to obtain meaningful data from visual inputs such as photos and videos. The insights gained from computer vision are then used to take automated actions. Just like AI gives computers the ability to ‘think’, computer vision allows them to ‘see’.



As humans, we generally spend our lives observing our surroundings using optic nerves, retinas, and the visual cortex. We gain context to differentiate between objects, gauge their distance from us and other objects, calculate their movement speed, and spot mistakes. Similarly, computer vision enables AI-powered machines to train themselves to carry out these very processes. These machines use a combination of cameras, algorithms, and data to do so.

2. AI enabled manipulation and grasping:

Artificial intelligence is about more than just enabling independent action–a robot in manufacturing can actually develop better ways to mechanically interact with the world. A robot using artificial intelligence in developing the best, most efficient ways to utilize its moving parts. Like most applications of AI in robotics, the bulk of work done in this area is done long before the robot is operating on the factory floor and is a part of an overall machine learning phase.
Grasping- machine learning and artificial intelligence provide direction to the robots with knowledge of the most powerful position to grasp an object.

3. AI enhanced navigation and motion control:

AI is improving robotic motion tasks by breaking down individual joint movements into motion primitives or sequences of movement. Advanced robotic arms have refined detection of unusual placement, friction, and gear slippage. For example, when motors and drives are in upper-level control, AI can respond to and manipulate uncommon changes in real time. This means that OEMs using AI will have the ability to perceive and make changes on devices to supply more torque, more current, and more feedback.

4. Real world perception:

- **Healthcare-**AI robotics is increasingly disrupting and transforming the healthcare market. ML-driven robotics is already a massive part of the healthcare chain, including function testing, surgery, research, data integration, etc.
- **Agriculture-**Integrating AI, ML, and Robotics provide agronomists with useful and actionable insights to help improve their farm productivity.
- **Warehouses-**Big companies with even larger warehouses are big consumers of robotics as it cuts operational time and intermediate costs.
- **Automobiles-**The role of robotics has a whole network of applications in the automotive industry ranging from designing, supply chain, and production activities to an entire set of management activities. Systems like driver assistance, autonomous driving, and driver risk assistance are being implemented in transportation for automobile industries.

5. Natural language processing:

Natural language processing (NLP) is a branch of artificial intelligence within computer science that focuses on helping computers to understand the way that humans write and speak. This is a difficult task because it involves a lot of unstructured data. The style in which people talk and write (sometimes referred to as 'tone of voice') is unique to individuals, and constantly evolving to reflect popular usage.

Understanding context is also an issue – something that requires semantic analysis for machine learning to get a handle on it. Natural language understanding (NLU) is a sub-branch of NLP and deals with these nuances via machine reading comprehension rather than simply understanding literal meanings. The aim of NLP and NLU is to help computers understand human language well enough that they can converse in a natural way.

Real-world applications and use cases of NLP include:

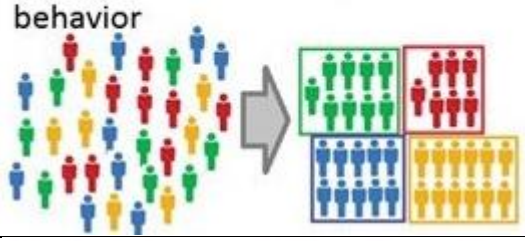
- Voice-controlled assistants like Siri and Alexa.
- Natural language generation for question answering by customer service chatbots.
- Streamlining the recruiting process on sites like LinkedIn by scanning through people's listed skills and experience.
- Tools like Grammarly which use NLP to help correct errors and make suggestions for simplifying complex writing.
- Language models like autocomplete which are trained to predict the next words in a text, based on what has already been typed.

Difference between Classification and Clustering:

Parameter	CLASSIFICATION	CLUSTERING
Type	used for supervised learning	used for unsupervised learning
Basic	process of classifying the input instances based on their corresponding class labels	grouping the instances based on their similarity without the help of class labels
Need	it has labels so there is need of training and testing dataset for verifying the model	there is no need of training and testing dataset

	created	
Complexity	more complex as compared to clustering	less complex as compared to classification
Example Algorithms	Logistic regression, Naive Bayes classifier, Support vector machines, etc.	k-means clustering algorithm, Fuzzy c-means clustering algorithm, Gaussian (EM) clustering algorithm, etc.
Need of training data set	It uses a training dataset.	It does not use a training dataset.

Difference between clustering and association

Clustering	Association
It splits the dataset into groups based on their similarities	allows you to establish associations amongst data objects inside large databases.
Clustering Grouping customers by purchasing behavior 	**Association** People that buy X tend to buy Y People that buy A+B tend to buy C 
Method of grouping the objects into clusters	It involve discovering patterns in data finding co-occurrences.
Discover inherent grouping in data	used for finding the relationships between variables in the large database
• K-means clustering • Hierarchical clustering • Density based clustering • EM clustering	• Market Basket Analysis

Difference between supervised and Unsupervised machine learning

Supervised Learning	Unsupervised Learning
Supervised learning algorithms are trained using labeled data.	Unsupervised learning algorithms are trained using unlabeled data.
Supervised learning model takes direct	Unsupervised learning model does not take

feedback to check if it is predicting correct output or not.	any feedback.
Supervised learning model predicts the output.	Unsupervised learning model finds the hidden patterns in data.
In supervised learning, input data is provided to the model along with the output.	In unsupervised learning, only input data is provided to the model.
The goal of supervised learning is to train the model so that it can predict the output when it is given new data.	The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset.
Supervised learning needs supervision to train the model.	Unsupervised learning does not need any supervision to train the model.
Supervised learning can be categorized in **Classification** and **Regression** problems.	Unsupervised Learning can be classified in **Clustering** and **Associations** problems.
Supervised learning can be used for those cases where we know the input as well as corresponding outputs.	Unsupervised learning can be used for those cases where 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.
Supervised learning is not close to true Artificial intelligence as in this, we first train the model for each data, and then only it can predict the correct output.	Unsupervised learning is more close to the true Artificial Intelligence as it learns similarly as a child learns daily routine things by his experiences.
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.