Unsupervised Learning in Machine Learning

Overview

Unsupervised learning is a type of machine learning where algorithms learn from unlabelled data. Unlike supervised learning, where the model is trained with labelled data, unsupervised learning identifies patterns and structures in the data without explicit guidance.

Key Concepts

1. Types of Unsupervised Learning

- **Clustering**: Grouping similar data points together.

- **Association**: Finding relationships between data points.

2. Clustering
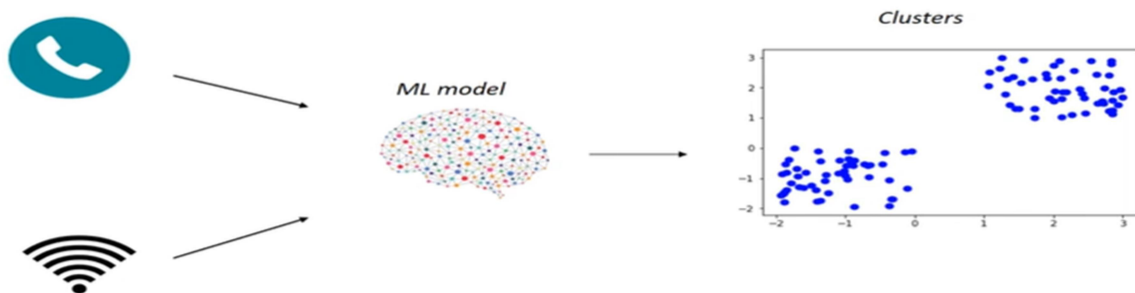
- **Definition**: Clustering involves grouping data points based on similarity. For example, if we have images of apples and mangoes, a clustering algorithm can automatically group these images based on their features without prior labelling.

- **Example**: A mobile network company can use clustering to analyse user data. By clustering users based on call duration and internet usage, the company can tailor offers to different user groups, enhancing revenue 2.

3. Association

Customer 1
- Bread
- Milk
- Fruits
- wheat

Customer 2
- Bread
- Milk
- Rice
- Butter

Customer 3

Now, when customer 3 goes and buys bread, it is highly likely that he will also buy milk.

- **Definition**: Association focuses on discovering interesting relationships between variables in large datasets. For instance, if customers frequently buy bread and milk together, an association algorithm can identify this pattern 3.

- **Example**: Supermarkets can use association rules to suggest products to customers. If a customer buys bread, they might also be prompted to buy milk, as these items are often purchased together 3.

Important Algorithms

- **Clustering Algorithms**:

  - **K-Means Clustering**: Partitions data into K distinct clusters based on feature similarity.

  - **Hierarchical Clustering**: Builds a hierarchy of clusters either agglomeratively or divisively.

- **Association Algorithms**:

  - **Apriori Algorithm**: Identifies frequent item sets and generates association rules.

  - **Eclat Algorithm**: A depth-first search algorithm for finding frequent item sets.

- **Dimensionality Reduction**:

  - **Principal Component Analysis (PCA)**: Reduces the dimensionality of data while preserving variance, useful for simplifying datasets with many features 4.

Questions to Consider

- What are the main differences between supervised and unsupervised learning?

- Can you explain how clustering algorithms can be applied in real-world scenarios?

- How does the Apriori algorithm work in finding associations in data?