11 Unsupervised Learning

- Unsupervised learning is a type of machine learning where the algorithm is trained on unlabeled data, meaning that the model is given input data without corresponding output labels
- The goal of unsupervised learning is to find hidden patterns, structures, or relationships in the data

Types of Unsupervised Learning

1. **Clustering**
    - Group similar data points into clusters or groups where data points within the same cluster are more similar to each other than to those in other clusters
        - **K-Means Clustering**
            - Partitions the data into a predetermined number of clusters by minimizing the variance within each cluster
        - **Hierarchical Clustering**
            - Builds a hierarchy of clusters either by starting with individual data points and merging them into larger clusters (agglomerative) or by starting with a single cluster and splitting it into smaller ones (divisive)
        - **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**
            - Forms clusters based on the density of data points, which allows it to find clusters of arbitrary shapes and handle noise (outliers)
    - **Gaussian Mixture Models (GMM)**:
        - Assumes that the data is generated from a mixture of several Gaussian distributions
        - Each cluster is represented by a Gaussian distribution, and the algorithm tries to find the parameters of these distributions
        - GMM is more flexible than K-means because it allows clusters to have different shapes and sizes
2. **Dimensionality Reduction**
    - Reduce the number of features or dimensions in the data while retaining as much of the important information as possible
        - **Principal Component Analysis (PCA)**
            - Transforms the data into a new set of orthogonal axes (principal components) that capture the maximum variance in the data
        - **t-Distributed Stochastic Neighbor Embedding (t-SNE)**
            - Reduces the dimensionality of data, particularly useful for visualizing high-dimensional data in 2 or 3 dimensions by preserving the local

            structure of the data
    - **Autoencoders**
        - Neural network models that learn to compress the input data into a lower-dimensional representation and then reconstruct it, useful for dimensionality reduction

3. **Association**
    - Discover interesting relationships or associations between different variables in the dataset
        - **Apriori Algorithm**
            - Finds frequent itemsets in transactional data and derives association rules, commonly used in market basket analysis
        - **Eclat Algorithm**
            - Similar to Apriori but uses a depth-first search approach to find frequent itemsets
        - **FP-Growth (Frequent Pattern Growth)**
            - An efficient algorithm that compresses the dataset using a structure called an FP-tree and finds frequent itemsets without candidate generation
4. **Anomaly Detection**
    - Identify rare or unusual data points that do not conform to the general pattern of the data
        - **Isolation Forest**
            - Detects anomalies by isolating data points in the feature space using random partitions
        - **One-Class SVM**
            - A variant of Support Vector Machine that tries to separate normal data from anomalies by learning a decision boundary around the normal data
        - **LOF (Local Outlier Factor)**
            - Identifies anomalies by comparing the local density of data points to that of their neighbors, where points with significantly lower density are considered anomalies

Applications of Unsupervised Learning

- **Market Basket Analysis**
    - Association rules help in discovering product associations for cross-selling
- **Customer Segmentation**
    - Clustering techniques group customers based on purchasing behavior, enabling targeted marketing
- **Anomaly Detection**

    - Identifying fraudulent activities or system failures
- **Data Compression**
    - Dimensionality reduction methods reduce the complexity of data, useful in image compression and noise reduction
- **Recommendation Systems**
    - Uncovering patterns in user preferences to suggest relevant content