TE - comps - A
9913
Mark Lopes

Comparison of classification and clustering algorithms

1. Introduction:

In the field of data science, machine learning algorithms can be broadly categorized into supervised and unsupervised learning techniques.

~~2. Classification~~
2. Overview of classification and clustering algorithms

2.1 Classification Algorithms

Classification is a supervised learning technique used for predicting the categorical class labels of new observations based on past observations.
Common algorithms include:-
i) Logistic regression
ii) Decision tree
iii) Support vector machine (SVM)
iv) Neural networks

2.2 Clustering Algorithms

Clustering is an unsupervised learning technique that groups similar data points together based on certain characteristics. Common clustering algorithms include:

i) K-mean clustering
ii) Hierarchical clustering
iii) DBSCAN (Density-based spatial clustering of applications with noise.

3. Comparison of classification and clustering algorithms.

3.1 Purpose and definition

i) Classification aims to assign data points to predefined categories or classes based on labeled training data

ii) Clustering seeks to group data points into clusters based on similarity, with no predefined labels.

4. Real - life case studies

4.1 Case study 1: classification in healthcare

i) Objective: To predict if a patient has a specific disease based on symptoms and historical data

ii) Algorithm: Decision trees and neural networks were applied to diagnose diabetes using a patient dataset.

iii) Results:

a. Accuracy: Neural networks achieved 92% accuracy, providing reliable predictions

b. Interpretability: Decision trees offered good interpretability allowing healthcare professionals to understand and trust the model.

iv) Impact: This model aids ~~both~~ healthcare providers in early disease diagnosis, leading to timely intervention and improved patient outcomes.

4.2 Case study 2: Clustering in retail customer segmentation.

i) Objective: To segment customers into distinct groups based on purchasing behaviours and demographics.

ii) Algorithm: K-means clustering was used to divide customers into five segments, identifying frequent buyers, occasional shoppers, discount seekers, etc.

iii) Results:

a) Interpretability: Clusters provided valuable insights, enabling ~~the~~ targeted marketing campaigns.

b) Scalability: K-means scaled effectively with over 100,000 customers, making it efficient for large scale analysis.

iv) Impact: By implementing targeted campaigns, customer engagement increased by 30% and marketing ~~for large scale analysis.~~ ROI improved.

5. Conclusion

Both classification and clustering algorithms have unique strengths and applications. Classification excels in scenarios where labeled data is available and prediction is the primary goal.

Clustering on the other hand, is invaluable for exploratory data analysis, particularly when understanding patterns within unlabeled data.