

Assignment: Clustering Vs Classification and Regression Vs Classification

Name: Apoorv Kashyap

INTERN ID: MST03-00

Differences Between Clustering and Classification

Clustering

1. **Definition:** Clustering is an unsupervised learning technique where the objective is to group a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups.
 - **Example:** Grouping customers based on purchasing behavior without predefined labels.
2. **Labels:** Clustering does not use predefined labels; it finds patterns and relationships in data to form clusters.
 - **Example:** Segmenting news articles into topics without knowing the topics in advance.
3. **Algorithm Types:** Common algorithms include K-Means, Hierarchical Clustering, and DBSCAN.
 - **Example:** Using K-Means to group images based on color similarity.
4. **Use Case:** Clustering is often used in exploratory data analysis to discover natural groupings in data.
 - **Example:** Identifying different species of plants in a botanical garden based on physical characteristics.
5. **Outcome:** The output of clustering is a set of clusters, where each cluster represents a group of similar data points.
 - **Example:** Grouping social media users based on their interaction patterns.

Classification

1. **Definition:** Classification is a supervised learning technique where the objective is to assign predefined labels to new observations based on a training dataset.
 - **Example:** Classifying emails as spam or not spam.

2. **Labels:** Classification uses predefined labels and trains the model to learn the mapping from input features to these labels.
 - **Example:** Diagnosing diseases from medical images where the labels are known diagnoses.
3. **Algorithm Types:** Common algorithms include Decision Trees, Random Forest, Support Vector Machines, and Neural Networks.
 - **Example:** Using a Random Forest to classify whether a tumor is benign or malignant.
4. **Use Case:** Classification is used when the categories are known and the goal is to categorize new data points into these predefined categories.
 - **Example:** Categorizing customer support tickets into different issue types.
5. **Outcome:** The output of classification is a label for each input data point, predicting the category it belongs to.
 - **Example:** Predicting the species of a flower based on its petal and sepal measurements.

Differences Between Regression and Classification

Regression

1. **Definition:** Regression is a supervised learning technique where the goal is to predict a continuous output based on input features.
 - **Example:** Predicting the price of a house based on its size, location, and other features.
2. **Output:** The output of regression is a continuous value, which can be any real number.
 - **Example:** Forecasting the temperature for the next week.
3. **Algorithm Types:** Common algorithms include Linear Regression, Polynomial Regression, and Support Vector Regression.
 - **Example:** Using Linear Regression to predict a student's final exam score based on study hours.
4. **Use Case:** Regression is used when the target variable is continuous and the goal is to predict a specific value.
 - **Example:** Estimating the sales revenue for the next quarter.
5. **Evaluation Metrics:** Common evaluation metrics for regression include Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.
 - **Example:** Evaluating the performance of a regression model predicting stock prices using MSE.

Classification

1. **Definition:** Classification is a supervised learning technique where the objective is to assign predefined labels to new observations based on a training dataset.
 - **Example:** Classifying images as either of cats or dogs.
2. **Output:** The output of classification is a discrete label from a set of predefined categories.
 - **Example:** Identifying whether a given email is spam or not spam.
3. **Algorithm Types:** Common algorithms include Logistic Regression, Decision Trees, Random Forest, and Naive Bayes.
 - **Example:** Using Logistic Regression to classify whether a patient has a certain disease based on medical test results.
4. **Use Case:** Classification is used when the target variable is categorical and the goal is to predict the category of new data points.
 - **Example:** Determining the genre of a song based on its audio features.
5. **Evaluation Metrics:** Common evaluation metrics for classification include Accuracy, Precision, Recall, F1 Score, and ROC-AUC.
 - **Example:** Measuring the performance of a classification model predicting loan defaults using Precision and Recall.