Clustering vs. Classification

Clustering

1. **Definition**: Clustering is an unsupervised learning technique used to group similar data points together based on their features.
    - **Example**: Grouping customers into segments based on their purchasing behavior.

2. **Purpose**: It identifies the inherent structure in the data without prior labels.
    - **Example**: Discovering natural groupings in social network data to identify communities.

3. **Algorithms**: Common algorithms include K-means, Hierarchical Clustering, and DBSCAN.
    - **Example**: Using K-means to segment an image into different regions.

4. **Output**: Produces clusters where each cluster contains data points that are more similar to each other than to those in other clusters.
    - **Example**: Segmenting plants in a dataset based on leaf measurements.

5. **Evaluation**: Quality is often evaluated using metrics like silhouette score, Davies-Bouldin index, or visually through cluster plots.
    - **Example**: Evaluating the compactness and separation of clusters in a plot of customer data.

Classification

6. **Definition**: Classification is a supervised learning technique used to assign predefined labels to new observations based on training data.
    - **Example**: Classifying emails as spam or non-spam.

7. **Purpose**: It predicts the category or class of new observations based on learned patterns from labeled data.
    - **Example**: Diagnosing medical conditions from patient symptoms.

8. **Algorithms**: Common algorithms include Decision Trees, SVM, Naive Bayes, and Neural Networks.
    - **Example**: Using a decision tree to classify species of flowers based on petal dimensions.

9. **Output**: Produces a model that can assign labels to new data points.
    - **Example**: Predicting if a loan application will be approved or rejected.
10. **Evaluation**: Performance is evaluated using metrics like accuracy, precision, recall, and F1 score.
    - **Example**: Using a confusion matrix to evaluate the performance of a model classifying handwritten digits.

Regression vs. Classification

Regression

11. **Definition**: Regression is a supervised learning technique used to predict a continuous outcome based on input features.
    - **Example**: Predicting house prices based on features like size, location, and age.

12. **Purpose**: It models the relationship between dependent and independent variables to make predictions about future data.
    - **Example**: Forecasting stock prices based on historical data.

13. **Algorithms**: Common algorithms include Linear Regression, Polynomial Regression, and Support Vector Regression.
    - **Example**: Using linear regression to predict the amount of rainfall based on past weather data.

14. **Output**: Produces a continuous value as the prediction.
    - **Example**: Predicting a student's score on a test based on hours studied.

15. **Evaluation**: Performance is evaluated using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared.

- **Example**: Evaluating a model predicting car prices by comparing predicted prices to actual prices.

Classification

16. **Definition**: Classification is a supervised learning technique used to assign predefined labels to new observations based on training data.
    - **Example**: Classifying emails as spam or non-spam.

17. **Purpose**: It predicts the category or class of new observations based on learned patterns from labeled data.
    - **Example**: Diagnosing medical conditions from patient symptoms.

18. **Algorithms**: Common algorithms include Decision Trees, SVM, Naive Bayes, and Neural Networks.
    - **Example**: Using a decision tree to classify species of flowers based on petal dimensions.

19. **Output**: Produces a model that can assign labels to new data points.
    - **Example**: Predicting if a loan application will be approved or rejected.

20. **Evaluation**: Performance is evaluated using metrics like accuracy, precision, recall, and F1 score.
    - **Example**: Using a confusion matrix to evaluate the performance of a model classifying handwritten digits.