**DIFFERENCE BETWEEN CLASSIFICATION AND CLUSTERING:**

**Learning Approach:**

* **Classification (Supervised Learning):** Classification is a supervised learning technique. This means the model learns from labeled data. The data consists of input features (attributes of the data point) and corresponding output labels (categories). The model analyzes this data to identify patterns and relationships between features and labels. Based on these learnings, it can then predict the labels for new, unseen data points.
* **Clustering (Unsupervised Learning):** In contrast, clustering is an unsupervised learning technique. The data used for clustering is unlabeled, meaning it doesn't have predefined categories. The goal of clustering is to group similar data points together based on their features, without any prior knowledge of the classes or categories that might exist.

**Objective:**

* **Classification:** The objective of classification is to assign a data point to a specific pre-existing class label. This helps in tasks like classifying emails as spam or not spam, or classifying images as containing cats, dogs, or birds.
* **Clustering:** The objective of clustering is to discover the inherent groupings (clusters) within the data. This helps in tasks like customer segmentation, where you might want to group customers with similar purchasing behavior, or anomaly detection, where you might want to identify data points that deviate significantly from the rest.

**Data Representation:**

* **Classification:** The data used for classification is typically well-defined and structured. Each data point has a set of features and a corresponding class label.
* **Clustering:** Clustering can work with various data formats, including structured data with numerical features, and unstructured data like text or images. The similarity between data points is determined based on the chosen distance or similarity metric.

**Evaluation:**

* **Classification:** The performance of a classification model is evaluated using metrics like accuracy, precision, recall, and F1 score. These metrics measure how well the model can correctly classify data points into their respective classes.
* **Clustering:** Evaluating clustering models is more subjective because there are no predefined class labels. Common approaches involve silhouette analysis, which measures the average distance between a data point and its closest points in a different cluster, or Calinski-Harabasz score, which compares the within-cluster variance to the between-cluster variance.

**Applications:**

* **Classification:** Classification has a wide range of applications across various domains, including spam filtering, image recognition, fraud detection, medical diagnosis, and customer segmentation.
* **Clustering:** Clustering applications include customer segmentation, market research (identifying groups with similar preferences), anomaly detection (identifying outliers in data), and image segmentation (grouping pixels with similar characteristics).

**In essence, classification predicts existing categories, while clustering discovers hidden categories.**

**DIFFERENCE BETWEEN REGRESSION VS CLASSIFICATION:**

**Target Variable:**

* **Regression:** Regression deals with predicting **continuous numerical values** as the target variable. These values can be any measurable quantity, like house prices, salary, temperature, or customer lifetime value.
* **Classification:** Classification, on the other hand, focuses on predicting **discrete categorical labels** for the target variable. These labels represent distinct classes or groups, like classifying emails as spam or not spam, images as containing cats or dogs, or customers as churned or active.

**Model Output:**

* **Regression:** A regression model outputs a continuous numerical value on a spectrum. For instance, predicting a house price of $350,000 or a temperature of 22 degrees Celsius.
* **Classification:** A classification model outputs a discrete class label. For example, classifying an email as "spam" or an image as containing a "dog."

**Task Type:**

* **Regression:** Regression is a **prediction** task. The model learns the relationship between input features and the continuous target variable to predict future values.
* **Classification:** Classification is a **classification** task. The model learns to distinguish between different categories based on the input features and assign data points to the most appropriate class.

**Evaluation Metrics:**

* **Regression:** Common metrics for evaluating regression models include **mean squared error (MSE)**, **root mean squared error (RMSE)**, and **mean absolute error (MAE)**. These metrics measure the difference between the predicted values and the actual values.
* **Classification:** Classification models are evaluated using metrics like **accuracy, precision, recall, and F1 score**. These metrics assess how well the model can correctly classify data points into their respective categories.

**Applications:**

* **Regression:** Regression is used in various applications where predicting continuous values is crucial. This includes tasks like stock price prediction, weather forecasting, demand forecasting, and real estate price estimation.
* **Classification:** Classification has a wide range of applications across various domains, including spam filtering, image recognition, fraud detection, medical diagnosis, and customer segmentation.