## Clustering vs. Classification

|  |  |  |
| --- | --- | --- |
| Feature | Clustering | Classification |
| **Learning Type** | Unsupervised | Supervised |
| **Data** | Unlabeled | Labeled (with pre-defined categories) |
| **Objective** | Identify groups (clusters) based on similarities | Predict category membership for new data |
| **Output** | Groups (clusters) of data points | Class labels for new data |
| **Complexity** | Generally less complex | Generally more complex |

**Real-Life Examples:**

* **Clustering:** A streaming service might use clustering to identify groups of users with similar viewing habits. Based on what users watch, the service can cluster viewers who enjoy comedies, documentaries, or action movies. This helps personalize recommendations for each user group.
* **Classification:** An email provider uses classification to filter spam. The system is trained on labeled emails (spam and not spam) and uses this knowledge to predict the category of a new incoming email.

## Regression vs. Classification

|  |  |  |
| --- | --- | --- |
| Feature | Regression | Classification |
| **Output Variable** | Continuous numerical value | Discrete category label |
| **Goal** | Predict a continuous value | Assign data points to predefined categories |
| **Example of Output** | Predicting house price based on size and location | Classifying an image as a cat or dog |
| **Data Used** | Can be continuous or discrete for independent variables | Independent variables can be continuous or discrete, but the dependent variable is categorical |
| **Evaluation Metrics** | Mean Squared Error (MSE), R-squared | Accuracy, Precision, Recall |

**Real-Life Examples:**

* **Regression:** A weather forecasting model might use regression to predict the temperature for tomorrow. The model is trained on historical data that includes temperature readings along with factors like humidity and pressure. Based on these factors, the model predicts a continuous temperature value for the next day.
* **Classification:** A social media platform might use classification to identify hate speech in comments. The system is trained on labeled comments (hate speech and non-hate speech) and uses this knowledge to predict whether a new comment belongs to the hate speech category.