Machine Learning Basics

1. Supervised Machine Learning

Supervised learning involves training a model on a labeled dataset, which means that each training instance is paired with an output label or target variable. The model's goal is to learn a mapping from inputs to the desired output. Typical supervised learning task is classification and regression.

Example:

Email Spam Detection:

Task: Classify emails as "spam" or "ham"

Dataset: A collection of emails where each email is labeled as either "spam" or "ham"

Algorithm: Classification Algorithms (e.g. Decision Trees, Random Forest, Support Vector Machine (SVM), Naïve Bayes Classifier, etc)

2. Unsupervised Machine Learning

Unsupervised learning involves training a model on data without labeled responses. The system tries to learn without a teacher. The goal is to identify patterns or groupings in the data. Typical unsupervised learning task involves clustring, Anomaly detection, dimensionality reduction.

Example:

Customer Segmentation:

Task: Group customers into segments based on purchasing behavior.

Dataset: A collection of customer data with features such as purchase history, age, location, etc.

Algorithm: K-Means Clustering, Hierarchical Cluster Analysis (HCA), etc.

3. Reinforcement Learning

Reinforcement learning involves training a learning system, called an agent, which involves observing the environment, make a sequence of decisions and perform action. Thus get rewards for good decisions and penalizing it for bad ones. The agent learns to maximize the cumulative reward over time.

Example:

Game Playing (Chess BOT):

Task: Develop an AI that can play chess at a high level.

Environment: The chessboard, pieces, and rules of chess.

Algorithm: Q-Learning, Deep Q-Networks (DQN), etc.

4. Classification VS Regression VS Clustering

- **Classification**: Classification is the task of classifying instances of the data into sub-categories, that are already provided by the target variable. Classification is part of supervised machine learning, where we have labeled data for training.

  Example: *Diabetes prediction model*. Where the target variable(Y) is 0 / 1. That means the patient may have diabetes or not. From the training data we have to build a ML model to predict whether the patient has diabetes(1) or not(0).

- **Regression**: Regression is the task of predicting real or continuous output value based on input data. Regression is part of supervised machine learning, where labeled data is provided.

  Example: *Solar Power prediction model*. Here, the target variable is continuous in nature. We need to predict real value that have least cost parameters.

- **Clustering**: Clustering is the action of grouping a set of objects in such a way that objects in the same group (cluster) are more similar to each other than to those in other groups.

  Example: *Customer segmentation* (grouping customers based on purchasing behavior).

Comparison:

Aspect	Classification	Regression	Clustering
Type of learning	Supervised	Supervised	Unsupervised
Algorithms used	Logistic Regression, Support Vector Machine(SVM), etc	Linear Regression, Random Forest Regressor, etc.	K-Means, Hierarchical Cluster Analysis (HCA)
Output	Categorical labels	Real value	Clusters (groups)
Example use case	Spam E-mail Detection	Stock Price Prediction	Anomaly Detection, Customer Segmentation