**Artificial intelligence**

Artificial Intelligence is the theory and development of computer systems trained to “think” and perform tasks that would normally require human intelligence.

**Machine learning**

Machine Learning (ML) is a subset of artificial intelligence that enables systems to learn from data and improve their performance on specific tasks without explicit programming. It identifies patterns in data and uses these patterns to make predictions or decisions. ML is widely applied in fields like image recognition, natural language processing, recommendation systems, and fraud detection.

**How machines learn?**

Machines learn by analysing data through algorithms. The process involves several steps:

* **Data Input**: Machines require high-quality data, such as text, images, or numbers, to analyse.
* **Algorithms**: Mathematical models identify patterns in the data. Different algorithms are suited for tasks like classification or regression.
* **Model Training**: The machine adjusts its internal parameters to minimize errors between predictions and actual outcomes.
* **Feedback Loop**: Predictions are compared with actual results, and errors are corrected iteratively.
* **Evaluation**: Models are tested on unseen data to ensure they generalize well to real-world scenarios.

A diagram of a machine learning

AI-generated content may be incorrect.

**1. Supervised learning**

Supervised learning is a machine learning technique where a model is trained using **labelled data**. Each input in the dataset is paired with a corresponding correct output, enabling the model to learn the relationship between inputs and outputs. The primary goal is to generalize this learning to make accurate predictions on new, unseen data.

**Key Characteristics of Supervised Learning**

* **Labelled Data**: The training dataset contains inputs with corresponding correct outputs (labels).
* **Learning Process**: The model learns by comparing its predictions to the actual labels and adjusting its parameters to reduce errors.
* **Prediction**: Once trained, the model can predict outputs for new, unseen inputs.

**Types of Supervised Learning**

**A. Regression**:

Regression in machine learning refers to a supervised learning technique where the goal is to predict a continuous numerical value based on one or more independent features. It finds relationships between variables so that predictions can be made. we have two types of variables present in regression:

* **Dependent Variable (Target)**: The variable we are trying to predict e.g. house price.
* **Independent Variables (Features)**: The input variables that influence the prediction e.g. locality, number of rooms.

**Types of Regression**

Regression can be classified into different types based on the number of predictor variables and the nature of the relationship between variables:

1. **Simple Linear Regression**  
   Models the relationship between one independent variable and one dependent variable using a straight line. It's useful for straightforward predictions like estimating house price based on size.
2. **Multiple Linear Regression**  
   Extends simple linear regression by incorporating multiple independent variables. It helps predict outcomes using several features, like house price based on size, location, and number of rooms.
3. **Polynomial Regression**  
   Captures non-linear relationships by adding polynomial terms to the model. Ideal for modelling curved trends, such as population growth or temperature changes over time.
4. **Ridge & Lasso Regression**  
   These are regularized linear models that prevent overfitting by penalizing large coefficients. Ridge uses L2 regularization, while Lasso uses L1, which can also perform feature selection.

**B. Classification**

Classification is a core task in supervised learning, where the goal is to predict discrete labels or categories for given input data. It involves training a model on a labeled dataset, where each input is associated with a specific output label. The model learns to map inputs to their corresponding labels, enabling it to classify new, unseen data accurately.  
**Types of classification problems** **in machine learning:**

**1. Binary Classification**

Classifies data into two categories.  
Example: Spam vs. Not Spam in email filtering.  
The model outputs either class A or class B.

**2. Multiclass Classification**

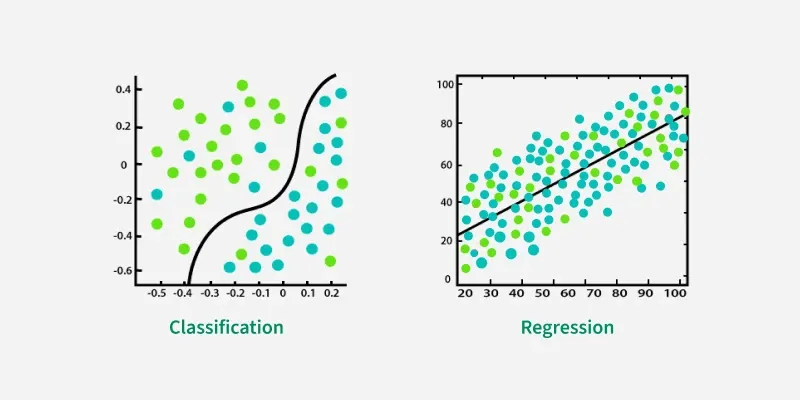
Classifies data into more than two categories, but only one label per instance.  
 Example: Classifying animals in images as either Cat, Dog, or Bird.  
Each input belongs to one of the available classes.

**3. Multi-label Classification**

Assigns multiple labels to a single instance.  
Example: Tagging a news article as both "Politics" and "Economy".  
Useful when data points can belong to multiple categories simultaneously.

**Common classification algorithms:**

1. **Logistic Regression**  
   A simple and effective algorithm for binary classification problems. It models the probability of a class using a logistic function.
2. **K-Nearest Neighbours (KNN)**  
   A non-parametric method that classifies data based on the majority class among its k-nearest neighbours in the feature space.
3. **Support Vector Machine (SVM)**  
   Finds the optimal hyperplane that separates classes with maximum margin. Works well for both linear and non-linear classification.
4. **Decision Tree**  
   Splits data into branches based on feature values, forming a tree structure. Easy to interpret and visualize.
5. **Random Forest**  
   An ensemble of decision trees that improves accuracy and reduces overfitting by averaging multiple trees.
6. **Naive Bayes**  
   Based on Bayes’ Theorem with the assumption of feature independence. Fast and effective for text classification and spam detection.
7. **Gradient Boosting (e.g., XGBoost, LightGBM)**  
   Builds models sequentially to correct errors of previous models. Highly accurate and widely used in competitions.
8. **Neural Networks**  
   Mimics the human brain structure with layers of interconnected nodes. Powerful for complex tasks like image and speech classification.



**2. Unsupervised learning**

In unsupervised learning, the algorithm is provided only with input data and no corresponding output labels. It autonomously analyses the data to group or organize it based on inherent patterns. For example, it might cluster customers based on purchasing behaviour or detect anomalies in network traffic.

**Key Techniques in Unsupervised Learning**

**A diagram of a learning process

AI-generated content may be incorrect.**

**A. Clustering**

Clustering involves grouping data points based on their similarities or differences. It is widely used in customer segmentation, image classification, and document clustering. Common clustering algorithms include:

* **K-Means Clustering**: Groups data into K clusters by minimizing the distance between points and their cluster centroids.
* **Hierarchical Clustering**: Builds a tree-like structure by iteratively merging or splitting clusters.
* **DBSCAN**: Identifies clusters in dense regions while treating scattered points as noise.
* **Mean-Shift Clustering**: Finds clusters by shifting points toward areas of higher density.

**B. Dimensionality Reduction**

Dimensionality reduction simplifies datasets by reducing the number of features while retaining essential information. This is useful for visualization and improving model performance. Common methods include:

* **Principal Component Analysis (PCA)**: Transforms data into uncorrelated components to reduce dimensions.
* **t-SNE**: Visualizes high-dimensional data in two or three dimensions.
* **Autoencoders**: Neural networks that compress and reconstruct data.

**C. Association Rule Learning**

This technique identifies relationships between variables in a dataset, often used in market basket analysis. For example, it can reveal that customers who buy milk are likely to buy bread. Popular algorithms include:

* **Apriori Algorithm**: Finds frequent itemsets and generates association rules.
* **FP-Growth Algorithm**: Efficiently identifies patterns without generating candidate sets.

**Applications of Unsupervised Learning**

Unsupervised learning has diverse applications across industries:

* **Customer Segmentation**: Groups customers based on purchasing behavior for targeted marketing.
* **Anomaly Detection**: Identifies unusual patterns for fraud detection or equipment failure.
* **Recommendation Systems**: Suggests products or content based on user behaviour.
* **Image and Text Clustering**: Organizes similar images or documents for classification or recommendation.
* **Medical Imaging**: Assists in diagnosing diseases through image segmentation and classification.

**3. Reinforcement Learning**

Reinforcement Learning (RL) is a branch of machine learning that focuses on how agents can learn to make decisions through trial and error to maximize cumulative rewards. RL allows machines to learn by interacting with an environment and receiving feedback based on their actions. This feedback comes in the form of rewards or penalties.

Reinforcement Learning revolves around the idea that an agent (the learner or decision-maker) interacts with an environment to achieve a goal. The agent performs actions and receives feedback to optimize its decision-making over time.

* **Agent**: The decision-maker that performs actions.
* **Environment**: The world or system in which the agent operates.
* **State**: The situation or condition the agent is currently in.
* **Action**: The possible moves or decisions the agent can make.
* **Reward**: The feedback or result from the environment based on the agent’s action.

**How Reinforcement Learning Works**

The RL process involves a loop where the agent:

1. Observes the current state of the environment.
2. Takes an action based on its policy (strategy).
3. Receives a reward or penalty and transitions to a new state.
4. Updates its policy to improve future decision-making.

A diagram of a agent and environment

AI-generated content may be incorrect.

**RL Algorithms**

* **Value-Based Methods:** Focus on learning the value of states or actions, such as Q-Learning, which uses the Bellman equation to update Q-values.
* **Policy-Based Methods:** Directly optimize the policy, such as the REINFORCE algorithm, which adjusts action probabilities to maximize rewards.
* **Model-Based Methods:** Use a model of the environment to simulate outcomes and plan actions, such as Model Predictive Control (MPC).

**Applications of Reinforcement Learning**

RL is widely used in:

* **Robotics**: Automating tasks and optimizing movements.
* **Game AI**: Developing strategies for games like chess and Go.
* **Industrial Control**: Real-time optimization of processes.
* **Personalized Training Systems**: Adapting content based on user behavior.