Types of Machine Learning

AIML-500

Introduction to Machine Learning

What is Machine Learning?

Definition: Machine Learning (ML) is a subset of artificial intelligence (AI) that enables systems to learn and improve from experience without being explicitly programmed.

Main Idea: ML uses data and algorithms to mimic the way humans learn, gradually improving its accuracy.

Types of Machine Learning

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Basic Comparison of ML Types

Supervised Learning: Trains on labeled data and learns to predict the output.

Unsupervised Learning: Works on unlabeled data and finds hidden structures.

Reinforcement Learning: Learns from interaction with the environment and focuses on decision-making.

Real-World Applications

Supervised Learning: Email filtering, disease prediction, fraud detection.

Unsupervised Learning: Recommendation systems, anomaly detection, gene sequence analysis.

Reinforcement Learning: Robotics, personalized advertising, self-driving cars.

Supervised Learning

- **Description:** In supervised learning, the algorithm is trained on labeled data, meaning the input and output are provided. The model learns to predict the output from the input.
- **Examples:** Classification (e.g., spam detection), Regression (e.g., predicting housing prices).
- **Common Algorithms:** Linear Regression, Decision Trees, Random Forests, Support Vector Machines (SVM).

Supervised Learning – Regression

Regression is used for predicting continuous numerical values. In this context, ML models attempt to learn the relationship between input features (variables) and a continuous output.

Example

Problem: Predicting house prices based on features like size, location, and age.
ML Task: A regression model (like linear regression or decision trees) would predict a specific number, such as a house price, given the input data.

The goal of regression is to find the function that best fits the relationship between inputs and outputs, minimizing the error between predicted and actual values.

Supervised Learning – Classification

Classification is used for predicting categorical outcomes (discrete labels or classes). The goal is to classify data points into predefined categories based on input features.

Example

Problem: Classifying whether an email is spam or not.
ML Task: A classification model (like logistic regression, decision trees, or support vector machines) assigns the email to one of two categories: "spam" or "not spam."

In classification, the goal is to find decision boundaries that separate classes and categorize new data points based on those boundaries.

Unsupervised Learning

- **Description:** In unsupervised learning, the algorithm is given data without explicit instructions on what to do. It tries to identify hidden patterns or structures in the data.
- **Examples:** Clustering (e.g., customer segmentation), Association (e.g., market basket analysis).
- **Common Algorithms:** K-Means Clustering, Hierarchical Clustering, Principal Component Analysis (PCA).

Unsupervised Learning - Clustering

Clustering

- **Purpose:** Group similar data points together based on their features.
- **Example:** Segmenting customers into different groups based on purchasing behavior for targeted marketing.
- **Algorithms:** K-means, hierarchical clustering, DBSCAN.

Unsupervised Learning - Dimensionality Reduction

Dimensionality Reduction

- **Purpose:** Reduce the number of features in a dataset while preserving important information.
- **Example:** Simplifying data for visualization or preprocessing data to improve the performance of other algorithms.
- **Algorithms:** Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE).

Unsupervised Learning - Anomaly Detection

Dimensionality Reduction

- **Purpose:** Identify data points that deviate significantly from the norm.
- **Example:** Detecting fraudulent transactions in financial data or identifying defective products in manufacturing.
- **Algorithms:** Isolation Forest, One-Class SVM, Local Outlier Factor (LOF).

Unsupervised Learning - Association Rule Learning

Dimensionality Reduction

- **Purpose:** Discover interesting relationships or associations between variables in large datasets.
- **Example:** Identifying frequent itemsets in market basket analysis, such as customers who buy bread also often buy butter.
- **Algorithms:** Apriori, Eclat.

Unsupervised vs. Supervised Learning

Aspect	Supervised Learning	Unsupervised Learning
Data Labels	Uses labeled data where the input features are paired with known outcomes (target values).	Uses unlabeled data where the outcomes are not known; the goal is to explore the structure of the data.
Goals	Focuses on predicting specific outcomes (continuous values for regression or discrete classes for classification).	Focuses on finding hidden patterns, grouping similar items, reducing dimensionality, or detecting anomalies without predefined categories.
Applications	Used for tasks where historical labeled data is available and specific predictions or classifications are needed.	Used for exploratory data analysis, pattern discovery, and understanding the underlying structure of data.

Reinforcement Learning

- **Description:** Reinforcement learning involves an agent that learns by interacting with an environment and receiving feedback in the form of rewards or penalties. The goal is to maximize the cumulative reward

- **Examples:** Game AI (e.g., AlphaGo), Robotics, Autonomous vehicles.

- **Common Algorithms:** Q-Learning, Deep Q-Networks (DQN), Policy Gradient Methods.

Challenges in Machine Learning

Data Quality: Garbage in, garbage out; models require clean, accurate data.

Bias and Fairness: Bias in training data can lead to unfair or unethical outcomes.

Scalability: As data grows, models need to scale efficiently.

Modeling Processes in Machine Learning

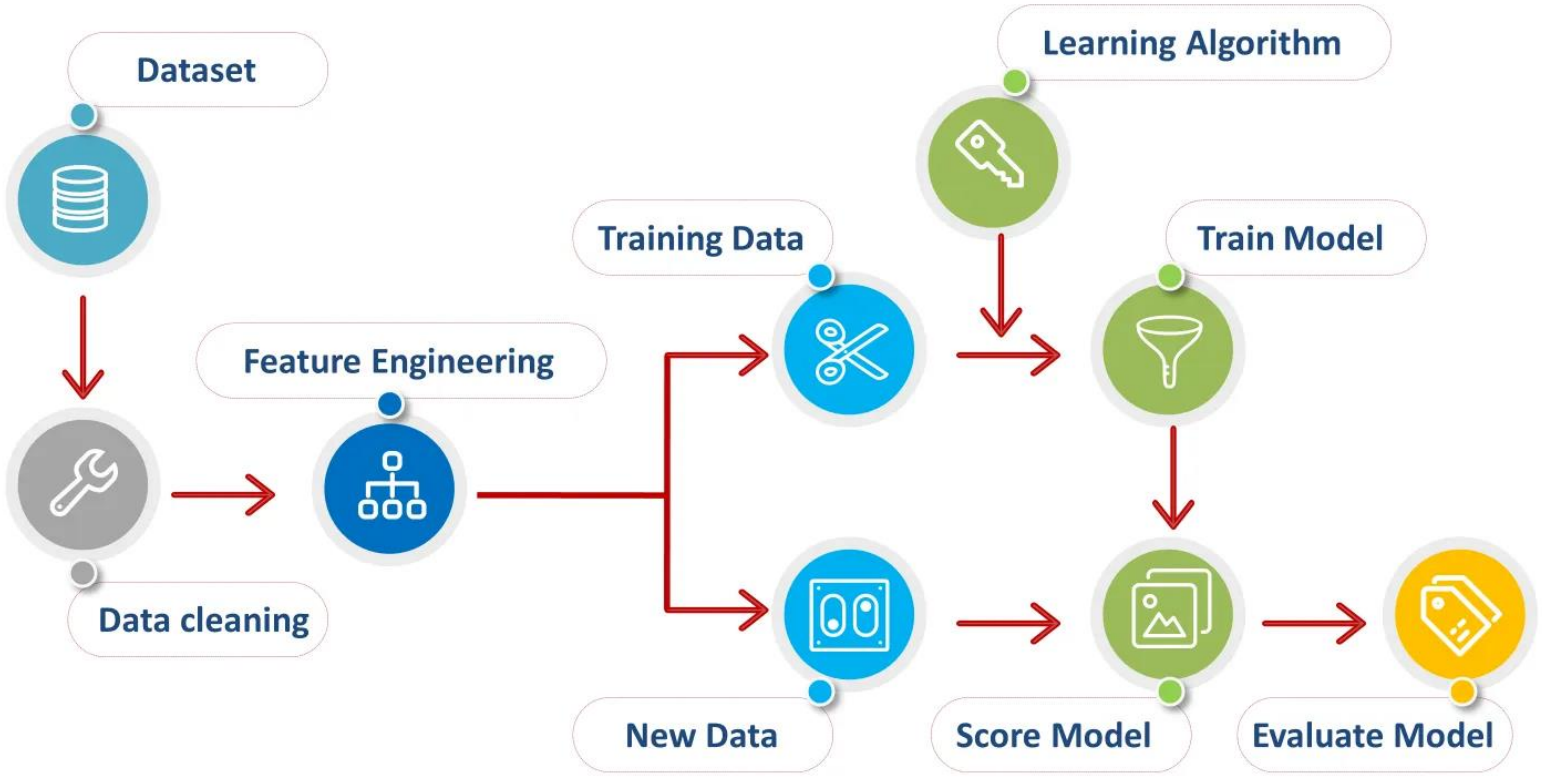
How can a model learn to predict an image label?

Source: VizWiz (2024) CC BY-4.0

Supervised Learning - Modeling Process

1. **Data Collection:** Gather labeled data.
2. **Data Preprocessing:** Clean and prepare the data.
3. **Model Selection:** Choose an appropriate algorithm (e.g., Linear Regression).
4. **Training:** Use the labeled data to train the model.
5. **Evaluation:** Test the model using unseen data and assess performance.
6. **Prediction:** Deploy the model for real-world predictions.

Supervised Learning – Example Flowchart



Example of Supervised Learning – Image Recognition

Convolutional Neural Network (CNN) for Image Classification.

Process Overview:

Input: Feed labeled images (e.g., cats and dogs) into the model.
Feature Extraction: The CNN extracts features (e.g., edges, shapes).
Prediction: The model predicts the label (e.g., cat or dog).

How can a model learn to cluster data?

Source: Lufthansa (2019)

Unsupervised Learning - Modeling Process

1. **Data Collection:** Gather labeled data.
2. **Data Preprocessing:** Standardize or normalize the data.
3. **Model Selection:** Choose a clustering or association algorithm (e.g., K-Means).
4. **Training:** The model identifies patterns or groupings in the data.
5. **Evaluation:** Use internal metrics (e.g., silhouette score) to assess clustering quality
6. **Application:** Use identified clusters or associations for insights.

Example of Unsupervised Learning – Market Basket Analysis

Apriori Algorithm for Association Rule Learning

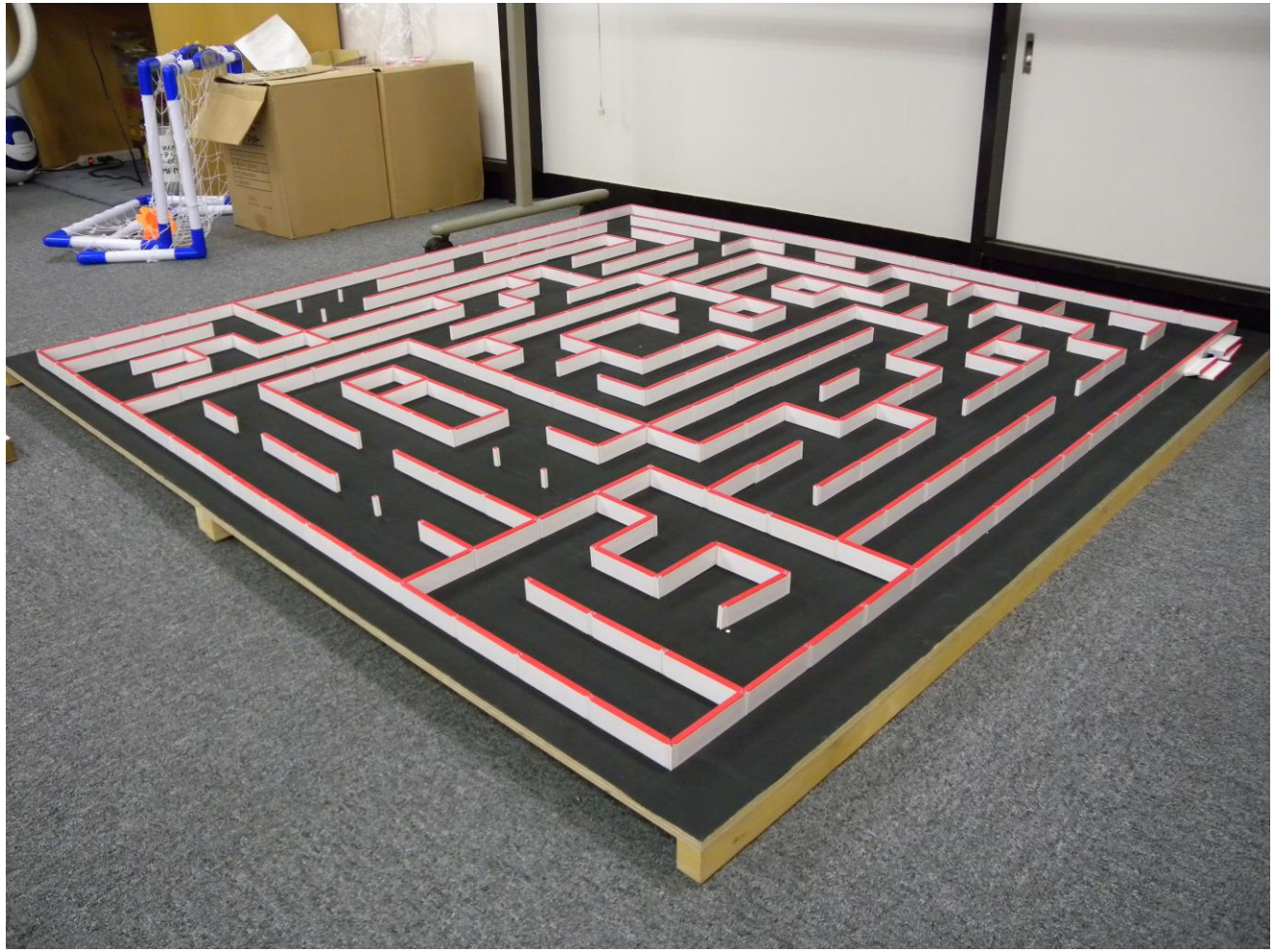
Process Overview:

Input: Transactional data (e.g., customer purchases).

Pattern Mining: The model finds frequent itemsets (e.g., milk and bread are often bought together).

Rule Generation: The model creates association rules (e.g., "If a customer buys milk, they are likely to buy bread").

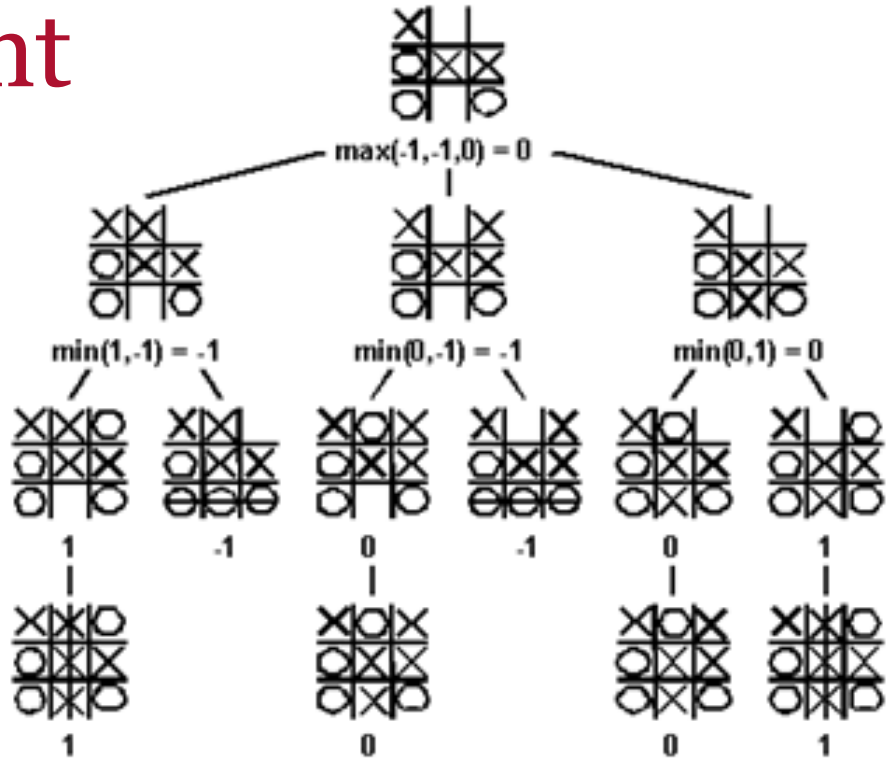
How does this robot learn to navigate the maze?



Reinforcement Learning - Modeling Process

1. **Agent Initialization:** Define the agent that will interact with the environment.
2. **Environment Setup:** Define the environment where the agent will operate.
3. **Reward System:** Establish rewards/penalties for different actions.
4. **Exploration & Exploitation:** The agent explores different actions and learns from outcomes.
5. **Policy Development:** The agent refines its policy to maximize cumulative rewards over time.
6. **Deployment:** The agent is applied in real-world tasks (e.g., robotics, gaming).

Example of Reinforcement Learning – Game AI



Q-Learning for Game AI
(e.g., learning to play
chess or tick-tac-toe)

Process Overview:
Input: The agent interacts with the game environment.
Exploration: The agent tries different moves (some good, some bad).
Reward: The agent receives points for correct moves and penalties for mistakes.
Improvement: The agent refines its strategy to maximize its score over time.

Learn More

All Machine Learning Models Explained in 5 Minutes | Types of ML Models Basics
https://youtu.be/yN7ypxC7838

AI system learns to play soccer from scratch
https://www.youtube.com/watch?v=foBwHVenxeU

A collection of machine learning interview materials
https://aiml.com/