

Introduction to Machine Learning

Introduction to Machine Learning

Quick?read, high?level overview for beginners

1. What Is Machine Learning (ML)?

Concept	Description
Definition	A sub?field of artificial intelligence that enables computers to learn patterns from data and make predictions or decisions without being explicitly programmed for each task.
Core Idea	Build a model that maps inputs ? outputs by extracting statistical relationships from examples.
Why It Matters	Automates complex tasks, handles large?scale data, adapts to changing environments, and powers many everyday technologies (recommendations, speech assistants, fraud detection, etc.).

2. Types of Learning Paradigms

Paradigm	Goal	Typical Data	Common Algorithms
Supervised Learning	Predict a known label (classification) or continuous value (regression).	Labeled examples (input?output pairs).	Linear/Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, Neural Networks.
Unsupervised Learning	Discover hidden structure without explicit labels.	Unlabeled data.	K?means, Hierarchical Clustering, DBSCAN, Principal Component Analysis (PCA), Autoencoders.
Semi?Supervised Learning	Leverage a small set of labeled data plus a large pool of unlabeled data.	Mixed labeled/unlabeled.	Self?training, Graph?based methods, Co?training.

Introduction to Machine Learning

Paradigm	Goal	Typical Data	Common Algorithms
Reinforcement Learning (RL)	Learn a policy that maximizes cumulative reward through interaction with an environment.	Sequential decision making episodes.	Q-learning, Deep Q-Network (DQN), Policy Gradient, Actor-Critic methods.
Self-Supervised Learning	Generate supervisory signals from the data itself (e.g., predicting masked parts).	Raw data (images, text, audio).	Contrastive learning, Masked language modeling (BERT).

3. Key Components of an ML Workflow

1. Problem Definition

- Clarify the objective (e.g., classify emails as spam/not-spam).
- Choose appropriate performance metric (accuracy, F1-score, RMSE, AUC, etc.).

2. Data Collection & Exploration

- Gather raw data from databases, APIs, sensors, etc.
- Perform exploratory data analysis (EDA): summary statistics, visualizations, outlier detection.

3. Data Preprocessing

- **Cleaning:** handle missing values, correct errors.
- **Transformation:** scaling/normalization, encoding categorical variables, feature engineering.
- **Splitting:** training / validation / test sets (e.g., 70/15/15) or cross-validation folds.

4. Model Selection

- Choose algorithm(s) that suit data size, feature type, interpretability needs.
- Consider baseline models (e.g., logistic regression) before complex ones.

5. Training

Introduction to Machine Learning

- Optimize model parameters on the training set (e.g., gradient descent).
- Use regularization (L1/L2, dropout) to prevent overfitting.

6. Evaluation

- Measure performance on validation data.
- Diagnose errors: confusion matrix, ROC curves, residual plots.

7. Hyperparameter Tuning

- Grid search, random search, Bayesian optimization, or automated tools (e.g., AutoML).

8. Testing & Deployment

- Final assessment on held-out test set.
- Package model (Docker, ON