

Machine Learning Basics – Study Resource

This document provides a concise introduction to Machine Learning concepts. It is designed to be used as reference material for learning and for training Retrieval-Augmented Generation (RAG) systems.

1. What is Machine Learning?

Machine Learning (ML) is a subset of Artificial Intelligence that enables systems to learn patterns from data and make predictions or decisions without being explicitly programmed.

2. Types of Machine Learning

- Supervised Learning: Models are trained using labeled data (e.g., Linear Regression, Decision Trees).
- Unsupervised Learning: Models discover patterns in unlabeled data (e.g., K-Means Clustering, PCA).
- Semi-Supervised Learning: Combines small labeled data with large unlabeled data.
- Reinforcement Learning: Agents learn by interacting with an environment and receiving rewards.

3. Common Machine Learning Algorithms

- Linear Regression – predicts continuous values.
- Logistic Regression – used for binary classification.
- Decision Trees – rule-based models for classification and regression.
- Random Forest – ensemble of decision trees.
- Support Vector Machines (SVM) – maximizes margin between classes.
- K-Nearest Neighbors (KNN) – classifies based on nearest data points.
- Neural Networks – inspired by biological neurons, used in deep learning.

4. Machine Learning Workflow

- Data Collection
- Data Preprocessing (cleaning, normalization, feature selection)
- Model Selection
- Model Training
- Model Evaluation
- Deployment and Monitoring

5. Evaluation Metrics

Common evaluation metrics include Accuracy, Precision, Recall, F1-Score, Mean Squared Error (MSE), and R-squared depending on the problem type.

6. Overfitting and Underfitting

Overfitting occurs when a model learns noise instead of patterns, performing well on training data but poorly on unseen data. Underfitting occurs when the model is too simple to capture underlying patterns.

7. Applications of Machine Learning

- Recommendation Systems
- Fraud Detection
- Image and Speech Recognition
- Natural Language Processing
- Predictive Analytics

End of Document. This resource can be used as training material for educational RAG-based evaluation systems.