Machine Learning Fundamentals

MACHINE LEARNING FUNDAMENTALS

Chapter 1: Introduction to Machine Learning

Machine Learning is a subset of Artificial Intelligence that enables systems to learn from data without being explicitly programmed.

1.1 Supervised Learning

In supervised learning, the algorithm learns from labeled data:

- Classification: Predicting discrete categories (e.g., spam/not spam)

- Regression: Predicting continuous values (e.g., house prices)

Common algorithms:

- Linear Regression

- Logistic Regression

- Decision Trees

- Random Forests

- Support Vector Machines (SVM)

- Neural Networks

1.2 Unsupervised Learning

Unsupervised learning finds patterns in unlabeled data:

- Clustering: Grouping similar data points

- Dimensionality Reduction: Reducing feature count while preserving information

Common algorithms:

- K-Means Clustering

- Hierarchical Clustering

- Principal Component Analysis (PCA)

- Autoencoders

1.3 Reinforcement Learning

Agents learn through trial and error, receiving rewards or penalties:

- Applications: Game playing, robotics, autonomous driving

- Key concepts: States, actions, rewards, policy

Chapter 2: Data Preparation

Quality data is crucial for ML success:

- Data Cleaning: Handling missing values and outliers

- Feature Engineering: Creating meaningful features from raw data

- Data Normalization: Scaling features to comparable ranges

- Train/Test/Validation Split: Ensuring proper model evaluation

Chapter 3: Model Evaluation

Metrics for assessing model performance:

- Classification: Accuracy, Precision, Recall, F1-Score, AUC-ROC

- Regression: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R^2 Score

Chapter 4: Common Pitfalls

- Overfitting: Model memorizes training data instead of learning general patterns

- Underfitting: Model too simple to capture data patterns

- Data Leakage: Information from test set influencing model training

- Imbalanced Datasets: Unequal class distributions affecting performance

Conclusion:

Successful machine learning requires careful planning, quality data, and proper evaluation methodologies.