

# Machine Learning - Quick Notes

## What is Machine Learning?

Machine Learning (ML) is a subset of AI that enables computers to learn from data and improve performance without explicit programming. Systems identify patterns and make decisions based on examples.

## Types of Machine Learning

### 1. Supervised Learning - Learns from labeled data (input-output pairs)

- Classification: Predicting categories (spam detection, disease diagnosis)
- Regression: Predicting continuous values (prices, temperatures)

### 2. Unsupervised Learning - Finds patterns in unlabeled data

- Clustering: Grouping similar data (customer segmentation)
- Dimensionality Reduction: Reducing features while preserving information

### 3. Reinforcement Learning - Learns through trial and error with rewards/penalties

- Applications: Game playing, robotics, autonomous vehicles

## Key Algorithms

- **Linear/Logistic Regression:** Simple prediction models
- **Decision Trees:** Tree-like decision models
- **Random Forest:** Ensemble of decision trees
- **Support Vector Machines (SVM):** Finds optimal class boundaries
- **K-Nearest Neighbors (KNN):** Classifies based on similarity
- **Neural Networks:** Layered networks for complex patterns
- **Deep Learning:** Multi-layer neural networks for images, text, speech

## ML Workflow

1. Define problem → 2. Collect data → 3. Preprocess data → 4. Engineer features → 5. Select model → 6. Train → 7. Evaluate → 8. Tune → 9. Deploy → 10. Monitor

## Key Metrics

**Classification:** Accuracy, Precision, Recall, F1-Score

## Common Challenges

- **Overfitting:** Model memorizes training data, fails on new data
- **Underfitting:** Model too simple to capture patterns
- **Data Quality:** Poor data = poor model
- **Imbalanced Data:** Unequal class representation

## Essential Concepts

- **Training Data:** Historical data for learning
- **Features:** Input variables for predictions
- **Labels:** Output values to predict
- **Model:** Mathematical representation learned from data

## Popular Tools

**Python Libraries:** Scikit-learn, TensorFlow, PyTorch, Pandas, NumPy

**Platforms:** Google Colab, Jupyter Notebooks, Kaggle

## Real-World Applications

Healthcare (diagnosis, imaging), Finance (fraud detection, trading), E-commerce (recommendations), Transportation (autonomous vehicles), Entertainment (Netflix, Spotify), NLP (chatbots, translation)

## Getting Started

1. Learn Python and statistics basics
2. Master Scikit-learn for classical ML
3. Practice on Kaggle datasets
4. Start simple, then progress to complex algorithms
5. Focus on understanding, not just coding