CRACKBITSWILP.IN

Detailed Study Guide: Machine Learning

This guide provides a comprehensive list of free online resources for each topic.

Machine Learning Course Topics and Study Resources

1. Introduction

1.1 Introduction to ML

- Machine Learning Tutorial (GeeksforGeeks) – A beginner-friendly overview defining ML as a branch of AI where systems learn from data, covering common types and examples
- Introduction to Machine Learning (Clarusway) – Explains ML fundamentals, how systems learn patterns from data without explicit programming
- Videos:
  - https://www.youtube.com/watch?v=6mSx_KJxcHI
  - https://www.youtube.com/watch?v=i_LwzRVP7bg

1.2 Objective of the Course

- Machine Learning Meaning & Objectives (Arramton blog) – Lists key ML objectives such as automating tasks, improving decision-making, and uncovering data insights.

1.3 Taxonomy (Types) of Machine Learning

- Types of Machine Learning (GeeksforGeeks) – Overview of supervised, unsupervised, and reinforcement learning (plus semi-supervised), with examples.

- Machine Learning Fundamentals (Clarusway) – Describes four primary ML strategies: supervised, unsupervised, semi-supervised, and reinforcement learning.

CRACKBITSWILP.IN

1.4 Design a Learning System

- Design a Learning System (GeeksforGeeks) – Tutorial on building an ML pipeline: data collection, preprocessing, feature selection, model training, and evaluation.
- Designing a Learning System (i2tutorials) – Steps for designing an ML system: defining training experience, target function, hypothesis space, and learning algorithm

1.5 Challenges in Machine Learning

- Challenges in ML: Issues and Solutions (IABAC) – Lists common ML challenges: poor-quality data, overfitting/underfitting, model bias, interpretability, data drift, etc.
- 5 Challenges in ML Adoption (MachineLearningMastery) – Highlights issues like data quality/availability, model complexity, cost/infra constraints, regulatory/ethical concerns, and skills gap

2. Machine Learning Workflow

2.1 Role of Data

- Role of Data in ML (Weskill Medium) – Emphasizes that data is crucial: "Without high-quality data, ML models are essentially useless," and model performance improves with better data quality/quantity.
- Understanding the Role of Data (lakeFS Blog) – Explains that "data is the core of any ML project," and that the volume and quality of data are key factors in training effective models

2.2 Data Preprocessing and Wrangling [IMPORTANT]

- What is Data Preprocessing? – Highlights that raw data often has errors, missing values, and noise; preprocessing cleans and transforms it into usable structured datasets for ML
- Data Preprocessing in ML (with python codes) – Notes that raw data is messy and that practitioners spend ~80% of their time on preprocessing to ensure high-quality input for models

CRACKBITSWILP.IN

2.3 Data Skewness Removal (Sampling)

- Handling Imbalanced Data – Explains that imbalanced (skewed) class distributions cause models to ignore minority classes and recommends sampling methods (oversampling/undersampling) to balance classes and improve performance
- How to remove data skewness – Describes oversampling the minority class (duplicating samples) and/or undersampling the majority class to overcome class imbalance and boost model accuracy

2.4 Model Training

- Model Training Tips (Ultralytics) – Continues training loop explanation: repeating the feed-loss-optimize steps over many epochs (iterations) so that the model's predictions become more accurate

2.5 Model Testing and Performance Metrics

- ML Performance Metrics (AltexSoft) – Explains that metrics (accuracy, precision, recall, F1, etc.) quantify a model's performance, answering "Is my model doing well?", which is critical for proper model testing

- Training vs Testing – Describes splitting data into training and testing sets: train the model on labeled training data, then evaluate it on unseen test data to measure how well it generalizes