Introduction to Machine Learning

How is Machine Learning different from Classical Programming?

Classical Programming has rigid rules written by programmers and a lot of hard coding (i.e., rules are predefined and do not change with data) is involved. At the same time, ML needs data as input and predicts based on the learned rules it learns after training on the data.

Classical Programming: Rigid rules written by programmers, a lot of hard coding.
Machine Learning (ML): Rules are learned from training a system/model on the data.

When to use ML over Classical Programming?

ML > Classical Prog. → applications that may not have easily visible patterns.

How do we categorize the different types of ML models?

Based on the type of Learning ML algo is categorized as:

- Supervised Learning: When output label (Y) is present in the data
- Unsupervised Learning: Output label (Y) is not in the data
- Reinforcement Learning: Used for AI in games, where there is a State, Environment, and Reward

Criteria: **Type of Learning**

1. **Supervised**: When output label (Y) is present in the training data
2. **Unsupervised**: Output label (Y) is not present in the training data
3. **Reinforcement**: Agent takes action in an “environment” for maximizing “reward” and changes its “state”

Criteria: **Type of Task**

1. **Classification**: Input data needs to be classified into categories
2. **Regression**: Input data needs to be mapped to a real-valued output
3. **Clustering**: Grouping of similar items together as one
4. **Recommendation**: To recommend items → more likable to the datapoint
5. **Forecasting**: Data pattern understanding → predicts future values

ML definition - ETP Framework

"Machine learning is the study of algorithms that improve their performance (P) at some task (T) with experience (E)."

Key Components

Task (T):

- Represents the specific problem or goal that the machine learning algorithm aims to solve.

Experience (E):

- It involves exposing the algorithm to data (training data) to learn patterns and relationships relevant to the task.

Performance (P):

- Quantitative measures are used to evaluate how well the model is performing the given task.

Examples:

- Stock Price Prediction:

  T: Predict the price of a stock.
  E: Use historical price data to train a model.
  P: Measure performance using a metric like Mean Squared Error.

- Customer Segmentation:

  T: Segment customers.
  E: Train a model using transactional data.
  P: Evaluate performance with a metric that maximizes inter-cluster dissimilarity and minimizes intra-cluster differences.

ML Examples

1. **Classification** - classify into one of the categories
    a. Example 1: Churn classification - whether an employee is likely to leave or not
    b. Example 2: Fraud Detection - whether a transaction is fraudulent

2. **Regression** - predict a real value
    a. Example 1: Predicting an individual's credit score based on income, debt, credit history, and other financial indicators.
    b. Example 2: Predicting horse price using sq area, location, and number of bedrooms

3. **Clustering** - group similar samples
    a. Example 1: Segmenting customer into various groups based on their spending habits, demographics, and preferences.
    b. Example 2: Group similar documents together based on their content or topics.

4. **Recommendation** - to recommend things
    a. Example 1: Netflix: Based on user interest and watch history, recommend a movie
    b. Example 2: E-commerce: Recommending products based on a user's purchase history, browsing behavior, and preferences.

5. **Forecasting** - predict the future value based on past patterns
    a. Example 1: forecast stock prices based on the last 10-day prices
    b. Example 2: Sales Forecasting: Predicting future sales volumes for products or services based on historical sales data, market trends, and seasonality.