Parul® University
Vadodara, Gujarat
NAAC GRADE A++

Information and Communication Technology

UNIT-1
Introduction to Machine Learning

Study Guide

Dr. Vinod Patidar
Associate Professor
CSE Department, PIT
Parul University

1. Introduction to Machine Learning

1.1 What is Machine Learning?

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that allows computers to learn pattern and make decisions from data without being explicitly programmed.

- **Definition (Tom Mitchell, 1997):**
  *"A computer program is said to learn from experience **E** with respect to some class of tasks **T** an performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**."*

Example:

- **Task (T):** Predict whether an email is spam or not.
- **Experience (E):** Observing labeled emails (spam/ham).
- **Performance (P):** Accuracy of correctly identifying spam.

Diagram: Basic ML Process

```
+--------------------+
| Input Data         |
+--------------------+
        |
        v
+--------------------+
| Learning Algorithm|
+--------------------+
        |
        v
+--------------------+
| Model (Hypothesis)|
+--------------------+
        |
```

```
          v
  +--------------------+
  |  Predictions/Output|
  +--------------------+
```

2. Learning Paradigms (Types of Machine Learning)

Machine Learning systems are categorized based on the nature of the “learning signal” or feedback available.

2.1 Supervised Learning

- The model learns from **labeled data** (input-output pairs).
- The goal is to learn a mapping f:X→Yf: X \rightarrow Yf:X→Y.

Example:

Predicting house prices based on size, location, etc.

Algorithm Examples:

- Linear Regression
- Decision Trees
- Support Vector Machines

- **Diagram:**

```
Input (X) ----> Model ----> Predicted Output (Y)
        ↑
        |
   True Output (Label)
```

2.2 Unsupervised Learning

- The model is given **unlabeled data** and must find hidden patterns or structure.

Example:

Grouping customers based on purchasing behavior.

Algorithm Examples:

- K-Means Clustering
- Principal Component Analysis (PCA)

2.3 Semi-Supervised Learning

- The dataset contains **a few labeled samples** and **many unlabeled samples**.
- Combines benefits of both supervised and unsupervised learning.

2.4 Reinforcement Learning

- The model learns by interacting with an environment and receiving **rewards or penalties**.

Example:

A robot learns to walk by trial and error.

Diagram:

```
Agent <---- Reward/Feedback ---- Environment
  |                        ^
  |                        |
  -------> Action/Decision ----------
```

2.5 Self-Supervised Learning (Modern Extension)

- Uses **data itself to create pseudo-labels**.
- Example: Predicting the missing part of an image or sentence (used in GPT, BERT).

3. PAC Learning (Probably Approximately Correct Learning)

3.1 Motivation

- Introduced by Leslie Valiant (1984).
- Provides a **theoretical foundation** for learning.
- Answers the question: *Can a hypothesis be learned from limited data and still generalize well?*

3.2 Intuition

PAC Learning formalizes how many samples are needed to ensure that the hypothesis we learn is **probably (with high probability)** and **approximately (within a small error)** correct.

3.3 Formal Definition

Let:

- H\mathcal{H}H = Hypothesis space
- DDD = Distribution of data
- fff = True target function
- ϵ\epsilonϵ = Maximum acceptable error (accuracy tolerance)
- δ\deltaδ = Confidence (probability that we're correct)

A hypothesis hhh is **PAC-learnable** if:

For every ϵ>0\epsilon > 0ϵ>0 and δ>0\delta > 0δ>0, there exists a polynomial number of samples mmm and algorithm such that with probability at least 1−δ1 - \delta1−δ,
the error of hhh on D is ≤ ε.

3.4 Diagram: PAC Concept

```
            True Error (ε)
         |<------------------->|
   +------------------------------------------+
   |                              |
   |       Hypothesis Space (H)              |
   |                              |
   +------------------------------------------+
              ↑
          Chosen Hypothesis (h)
    Probably (1 - δ) within ε error margin
```

3.5 Example

Suppose a spam filter predicts whether an email is spam.

- ϵ=0.05\epsilon = 0.05ϵ=0.05 → error < 5% acceptable
- δ=0.1\delta = 0.1δ=0.1 → 90% confidence

We want the algorithm to output a model that, with **90% confidence**, misclassifies **no more than 5%** of emails.

4. Basics of Probability in ML

4.1 Why Probability?

Machine Learning often deals with **uncertainty** — data may be noisy or incomplete. Probability helps quantify this uncertainty.

4.2 Key Concepts

Concept	Formula	Meaning
Probability of Event A	P(A)=favorable outcomestotal outcomesP(A) = \frac{\text{favorable outcomes}}{\text{total outcomes}}P(A)=total outcomesfavorable outcomes	Likelihood of A happening
Joint Probability	(P(A, B) = P(A) \times P(B	A))
Conditional Probability	(P(A	B) = \frac{P(A,B)}{P(B)})
Bayes' Theorem	(P(A	B) = \frac{P(B

4.3 Example

Suppose 1% of emails are spam and 90% of spam emails contain the word "offer."

Find the probability that an email with "offer" is spam:

P(Spam|Offer)=P(Offer|Spam)P(Spam)P(Offer)P(\text{Spam} | \text{Offer}) = \frac{P(\text{Offer}|\text{Spam})P(\text{Spam})}{P(\text{Offer})}P(Spam|Offer)=P(Offer)P(Offer|Spam)P(Spam)

This forms the base for **Naïve Bayes Classifier**.

4.4 Diagram: Conditional Probability

```
+------------------------+
|       P(A ∩ B)        |
+----------+-------------+
           |
           v
P(A|B) = P(A ∩ B) / P(B)
```

5. Version Spaces

5.1 Concept

A **Version Space (VS)** is the set of all hypotheses consistent with the observed training data.

Introduced by **Tom Mitchell**, it helps visualize how learning **reduces uncertainty** as more examples are observed.

5.2 Definition

Given:

- Hypothesis space HHH
- Training examples DDD

Then,

VSH,D={h∈H | h(xi)=yi for all (xi,yi)∈D}VS_{H,D} = \{ h \in H \ | \ h(x_i) = y_i \text{ for all } (x_i, y_i) \in D \}VSH,D={h∈H | h(xi)=yi for all (xi,yi)∈D}

i.e., the set of all hypotheses that correctly classify all training examples.

5.3 Candidate Elimination Algorithm

- Maintain two sets:
  - **S (Specific boundary):** most specific hypotheses.
  - **G (General boundary):** most general hypotheses.
- With each new example:
  - **Positive example:** generalize S, specialize G.
  - **Negative example:** specialize S, generalize G.

5.4 Diagram: Version Space

Hypothesis Space (H)

```
+----------------------------------------------+
|      G (General Boundary)                    |
|    +--------------------------------+        |
|    |     Version Space (VS)         |        |
|    |  +---------------------+       |        |
|    |  | S (Specific Boundary)|      |        |
|    |  +---------------------+       |        |
|    +--------------------------------+        |
+----------------------------------------------+
```

5.5 Example

Task: Identify “fruit is apple” based on attributes:

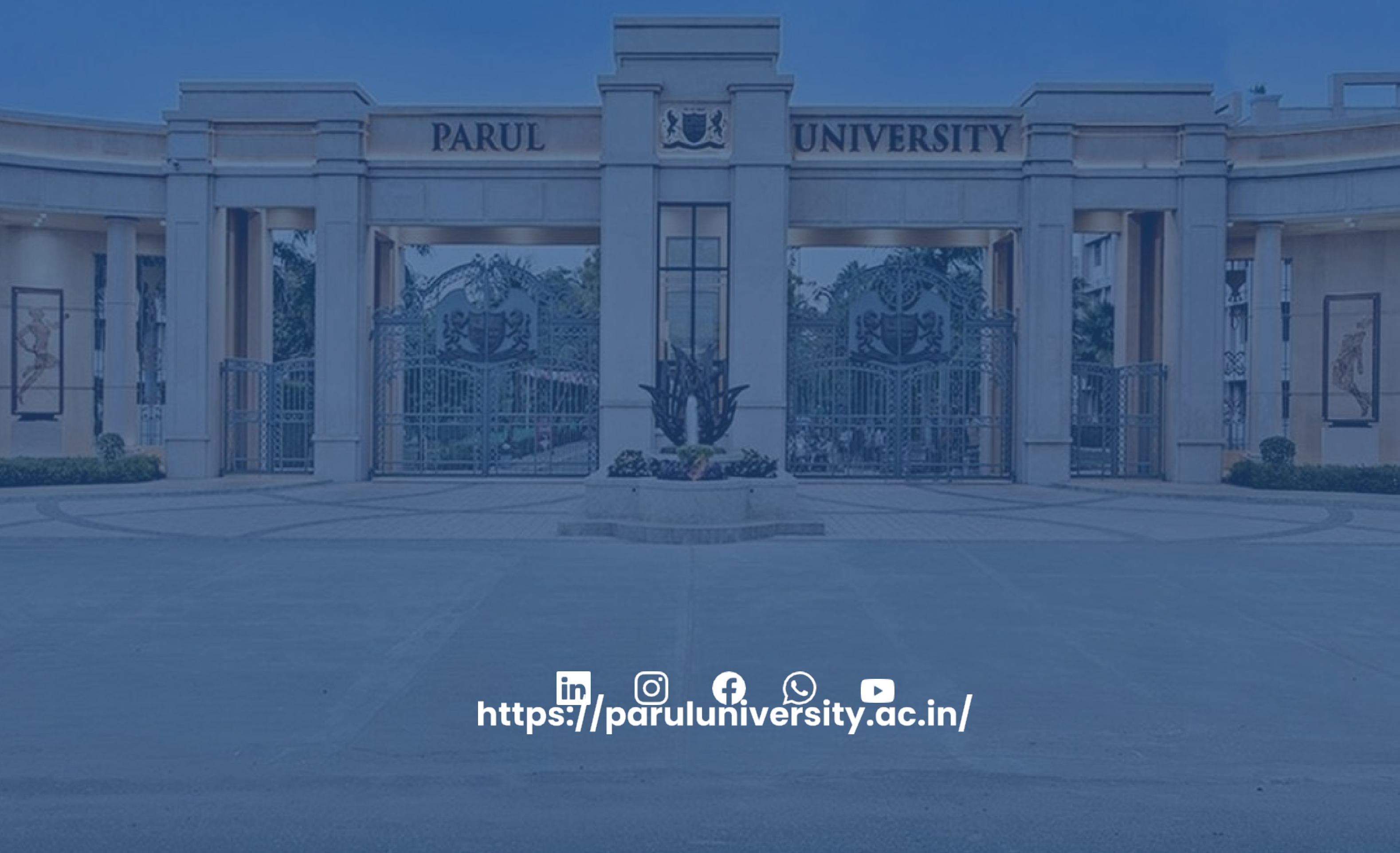
- Color (Red/Green)
- Size (Small/Large)

Example	Color	Size	Class
1	Red	Small	✅ Apple
2	Green	Large	❌ Not Apple

- After example 1:
    - S={(Red,Small)}S = \{(Red, Small)\}S={(Red,Small)}
    - G={(?,?)}G = \{(?, ?)\}G={(?,?)}
- After example 2:
    - S={(Red,Small)}S = \{(Red, Small)\}S={(Red,Small)}
    - G={(Red,?),(?,Small)}G = \{(Red, ?), (?, Small)\}G={(Red,?),(?,Small)}

Parul® University
Vadodara, Gujarat

NAAC GRADE A++



https://paruluniversity.ac.in/