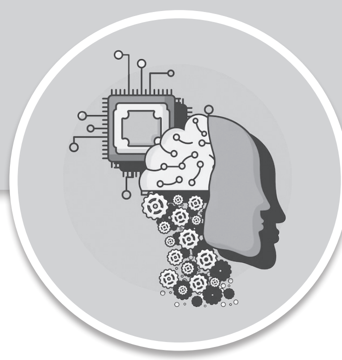
DATA SCIENCE & ARTIFICIAL INTELLIGENCE

Machine Learning



Comprehensive Theory
with Solved Examples *and* Practice Questions

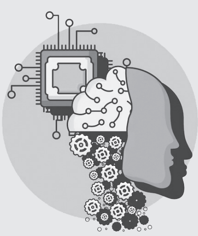
MADE EASY PUBLICATIONS

www.madeeasypublications.org

MADE EASY Publications Pvt. Ltd.
Corporate Office: 44-A/4, Kalu Sarai (Near Hauz Khas Metro Station), New Delhi-110016 | **Ph. :** 9021300500
Email : infomep@madeeasy.in | **Web :** www.madeeasypublications.org

Machine Learning

© Copyright by MADE EASY Publications Pvt. Ltd.
All rights are reserved. No part of this publication may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photo-copying, recording or otherwise), without the prior written permission of the above mentioned publisher of this book.



MADE EASY Publications Pvt. Ltd. has taken due care in collecting the data and providing the solutions, before publishing this book. Inspite of this, if any inaccuracy or printing error occurs then **MADE EASY Publications Pvt. Ltd.** owes no responsibility. We will be grateful if you could point out any such error. Your suggestions will be appreciated.

EDITIONS

First Edition : 2025

Second Edition : 2026

CONTENTS

Machine Learning

CHAPTER 1

Introduction to Machine Learning

1.1 WHAT IS MACHINE LEARNING?

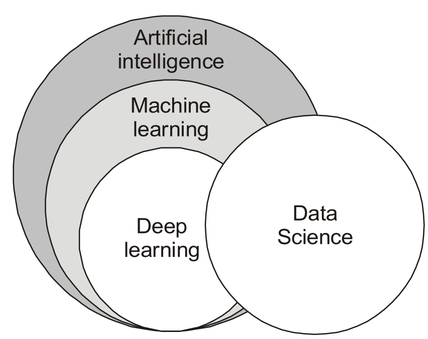
1.1.1 Definition and Overview

Understanding the Concept of Machine Learning

Machine learning is like teaching a computer to learn from examples instead of giving it specific instructions for every task. Imagine you have a box of Lego bricks and you want to build a specific model, like a car. Instead of following a step-by-step instruction manual, you observe many different Lego cars and learn how they are put together. After looking at enough examples, you can build your own Lego car even if you don't have the instructions.

In machine learning:

- Learning from Data: The computer looks at lots of examples (data) and figures out how to do a task by itself. For example, if you give it lots of pictures of cats and dogs, it can learn to tell the difference between them.
- Automation and Adaptation: Once it learns from the examples, it can make predictions or decisions on new data. For instance, if it sees a new picture of an animal, it can guess whether it's a cat or a dog.



Differentiating Machine Learning from Traditional Programming

Traditional programming is like following a recipe to bake a cake. You have specific instructions to follow, and you get the same result every time. In traditional programming:

1. Input Data: Ingredients like flour, sugar, eggs.
2. Program: The recipe with step-by-step instructions.
3. Output: The baked cake.

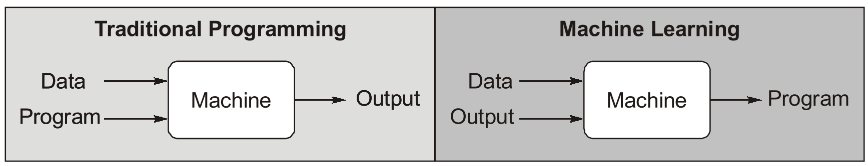
Machine learning is like teaching someone to bake a cake by letting them taste many different cakes and learn what good cakes taste like. Over time, they learn to bake a delicious cake without needing a precise recipe.

In machine learning:

1. Input Data: Lots of different cakes and their recipes (data).
2. Learning Algorithm: The person tastes and learns the patterns in good cakes (algorithm).
3. Model: The person's understanding of how to bake a cake (model).
4. Output: A new, delicious cake based on what they learned.

Key differences:

- Explicit Programming vs. Data-Driven Learning: Traditional programming uses fixed instructions, while machine learning uses data to create flexible rules.
- Static vs. Dynamic: Traditional programs don't change unless reprogrammed. Machine learning models adapt and improve with more data.



1.1.2 Historical Context and Evolution

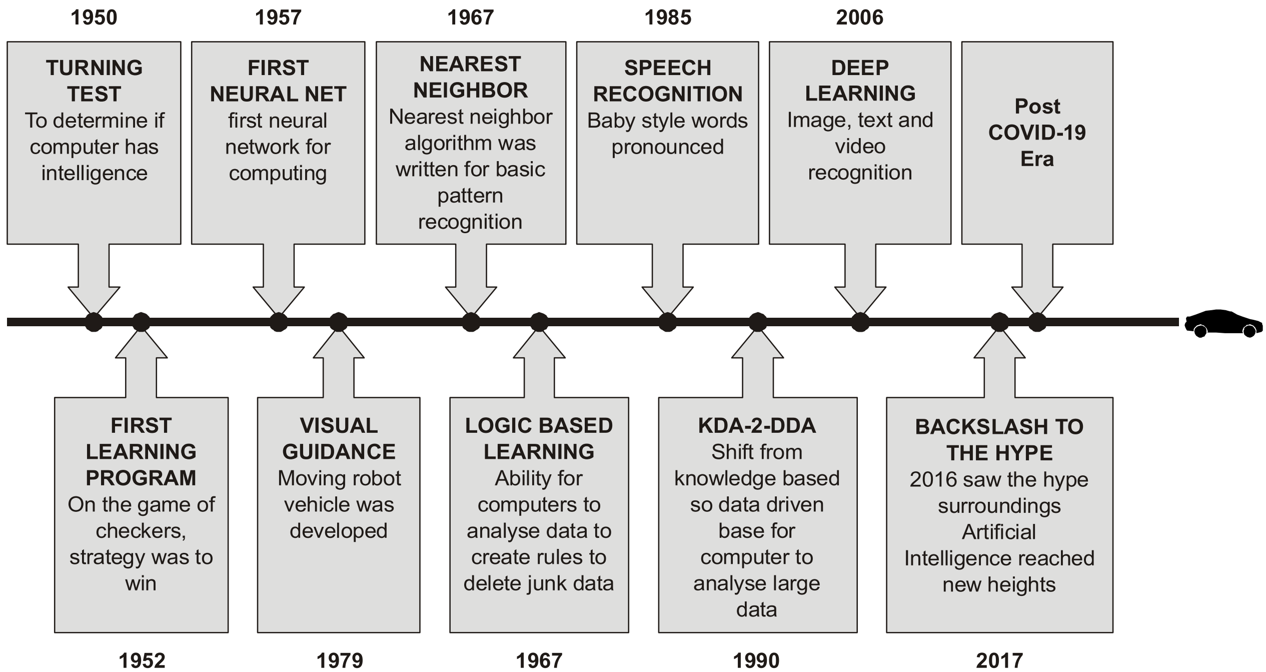
Early Developments in Machine Learning

The journey of machine learning began when scientists and engineers started wondering if machines could learn from data. Here are some key moments:

- 1950s: The term "machine learning" was coined by Arthur Samuel. He created a program that could play the game of checkers. By playing many games and learning from each move, the program got better over time.
- Perceptron: In 1957, Frank Rosenblatt introduced the perceptron, an early type of artificial neural network, which was designed to recognize patterns and learn from data.

Key Milestones and Breakthroughs

1. 1960s-1970s: Foundational algorithms were developed, such as decision trees, which help make decisions by splitting data into branches, and the *k*-nearest neighbors algorithm, which makes predictions based on the closest examples in the data.
2. 1980s: Neural networks became more sophisticated with the introduction of the backpropagation algorithm, allowing for more efficient training of multi-layer networks, similar to how our brain processes information.
3. 1990s: New techniques like Support Vector Machines (SVM) and ensemble methods (combining multiple models for better performance) emerged, improving the accuracy and robustness of machine learning models.
4. 2000s: Advances in computer hardware, especially Graphics Processing Units (GPUs), enabled the training of larger and more complex models, leading to the resurgence of neural networks and deep learning.
5. 2010s: With the explosion of big data and more powerful algorithms, machine learning achieved significant breakthroughs in image and speech recognition, making it possible for machines to understand and interpret complex data like humans.



Key contributions:

- Statistical Learning: Combining statistics with machine learning helped improve model accuracy by better understanding data patterns.
- Computational Power: Faster and more powerful computers allowed for the training of complex models on large datasets.
- Interdisciplinary Research: Collaboration across fields like computer science, statistics, and neuroscience led to innovative algorithms and applications.

Example 1.1 Which of the following is considered the first instance of a machine learning program?

(a) Backpropagation algorithm in neural networks
(b) Arthur Samuel's checkers-playing program
(c) Support Vector Machines (SVM)
(d) K-means clustering algorithm

Solution:(b)

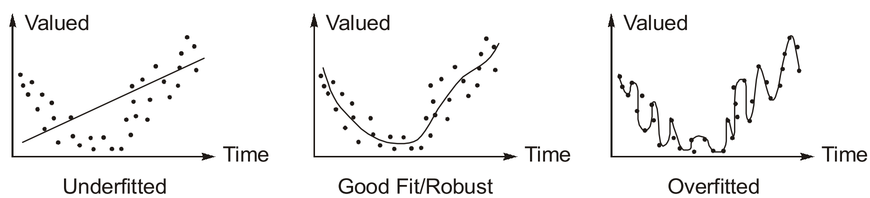
In the 1950s, Arthur Samuel created a checkers-playing program that improved its performance by learning from previous games, marking one of the earliest instances of machine learning.

1.1.3 Key Concepts and Terminology

Definitions of Basic Terms

- Algorithm: A set of rules or instructions the computer follows to solve a problem. In machine learning, algorithms process data to build models. Think of it as a recipe that tells the computer how to learn from data.

- **Model:** The outcome of a machine learning algorithm. It's a mathematical representation of a real-world process. For example, a model might predict house prices based on features like size, location, and number of bedrooms.
- **Training Data:** The examples the algorithm learns from. If you're teaching a computer to recognize fruits, the training data would include many pictures of fruits labeled with their names.
- **Features:** The individual pieces of data used to make predictions. In a dataset of house prices, features might include the number of rooms, size of the house, and location.
- **Labels:** The correct answers provided in the training data. For fruit recognition, the labels are the names of the fruits in the pictures.
- **Overfitting:** When a model learns the training data too well, including the noise, and performs poorly on new data. It's like memorizing answers to specific questions instead of understanding the concepts.
- **Underfitting:** When a model doesn't learn enough from the training data and performs poorly on both training and new data. It's like having a vague understanding and making incorrect predictions.



Introduction to the Learning Process

- **Training:** The algorithm uses training data to create a model. For example, it looks at pictures of cats and dogs (training data) and learns their features to build a model that can tell them apart.
- **Validation:** A step to tune model parameters and select the best model. Validation data helps to prevent overfitting. Think of it as a practice test to see how well the model is learning.
- **Testing:** Evaluating the model's performance on a separate, unseen dataset (test data) to assess how well it generalizes to new data. It's like a final exam to test the model's understanding.

Example to Illustrate Concepts

Imagine you want to teach a computer to recognize different types of fruit. Here's how the machine learning process would look:

1. **Collect Data:** Gather lots of pictures of apples, oranges, and bananas, and label each picture with the correct fruit name.
2. **Features and Labels:** The features are the characteristics of the pictures, like color, shape, and size. The labels are the names of the fruits.
3. **Training:** Feed the pictures (training data) into a machine learning algorithm. The algorithm learns the patterns and builds a model.
4. **Validation:** Use a separate set of pictures (validation data) to check how well the model is learning and make adjustments if needed.
5. **Testing:** Finally, test the model with new pictures it hasn't seen before to see how accurately it can identify the fruits.

This process helps the computer learn to recognize fruits from pictures without needing explicit instructions for each type.

Student's Assignments

Q.1 Which decade saw the development of the backpropagation algorithm, leading to improvements in neural networks?
(a) 1960s
(b) 1980s
(c) 1990s
(d) 2010s

Q.2 What was a key advancement in the 2000s that helped machine learning grow significantly?
(a) Introduction of decision trees
(b) Statistical learning theories
(c) Advancements in GPU computing
(d) The invention of deep learning

Q.3 Which factor contributed to major breakthroughs in machine learning during the 2010s?
(a) The invention of ensemble methods
(b) The availability of big data and powerful algorithms
(c) The introduction of decision trees
(d) The rise of perceptron-based models

Q.4 Which of the following best describes an algorithm in machine learning?
(a) A trained model used to make predictions
(b) A dataset containing labeled examples
(c) A set of rules or instructions for processing data and building models
(d) The final output of a machine learning model

Q.5 What is the training data in a machine learning model?
(a) The correct answers the model predicts after training
(b) The examples the algorithm learns from
(c) The mathematical representation of a real-world process
(d) The final predictions made by the model

Q.6 What is underfitting in machine learning?
(a) When a model fails to learn enough from the training data and performs poorly on both training and test data
(b) When a model memorizes training data and fails to generalize
(c) When a model predicts perfectly on training data but fails on test data
(d) When a model has too many features leading to high accuracy

Q.7 Which of the following is an example of a classification task?
(a) Predicting the price of a house based on size and location
(b) Grouping customers into different clusters based on purchasing behavior
(c) Detecting whether an email is "spam" or "not spam"
(d) Identifying common patterns in large datasets without predefined labels

Q.8 In a regression task, what type of output does the model predict?
(a) A discrete category, such as "dog" or "cat"
(b) A continuous value, such as house price or temperature
(c) A group or cluster without predefined labels
(d) A random guess based on probability

Q.9 Which machine learning task is an example of unsupervised learning?
(a) Predicting stock prices based on historical data
(b) Classifying emails as spam or not spam
(c) Grouping documents based on topic similarity
(d) Identifying handwritten digits with labeled examples

Q.10 When should semi-supervised learning be used?
(a) When labeled data is abundant and easy to obtain
(b) When no labeled data is available at all
(c) When there is a small set of labeled data and a large set of unlabeled data
(d) When supervised learning models perform perfectly

Answers Keys

1. (b)	**2.** (c)	**3.** (b)	**4.** (c)	**5.** (b)
6. (a)	**7.** (c)	**8.** (b)	**9.** (c)	**10.** (c)

Explanations

1. (b)

The backpropagation algorithm, which significantly enhanced the training of multi-layer neural networks, was introduced in the 1980s.

2. (c)

In the 2000s, improvements in Graphics Processing Units (GPUs) allowed for efficient training of larger and more complex machine learning models, fueling advancements in deep learning.

3. (b)

In the 2010s, machine learning achieved major progress in image and speech recognition due to the availability of large datasets and improved algorithms.

4. (c)

An algorithm in machine learning is a set of rules or instructions that process data to build a model. It acts as a recipe guiding the learning process.

5. (b)

Training data consists of labeled examples that the algorithm uses to learn patterns and relationships, helping it build an accurate model.

6. (a)

Underfitting occurs when a model doesn't learn sufficient patterns from the training data, leading to inaccurate predictions on both training and new data.

7. (c)

Classification assigns discrete labels to data, such as classifying emails as “spam” or “not spam”.

8. (b)

Regression predicts continuous values based on input features, such as estimating house prices.

9. (c)

Clustering documents based on shared characteristics is an example of unsupervised learning.

10. (c)

Semi-supervised learning is beneficial when labeling data is difficult, but there is plenty of unlabeled data to learn from.