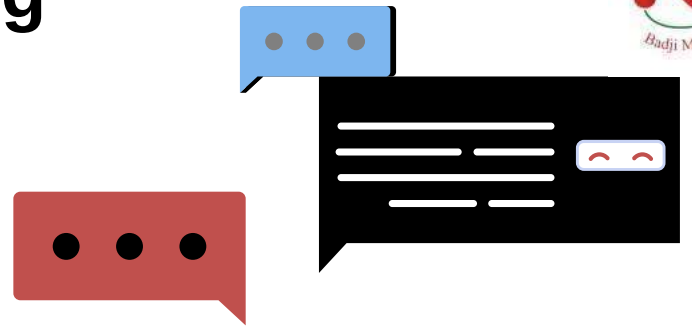
Chapter 4 : Machine learning



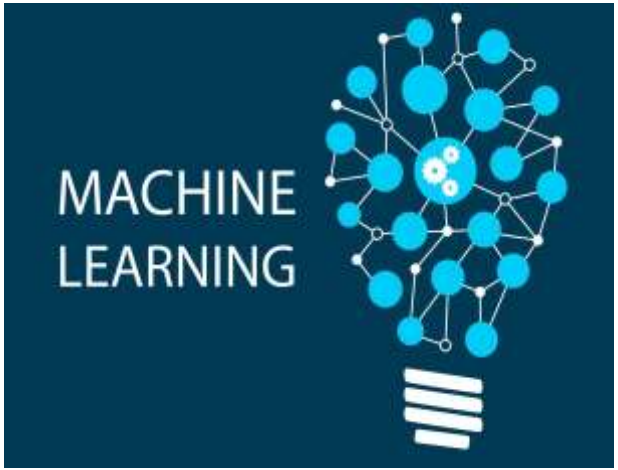


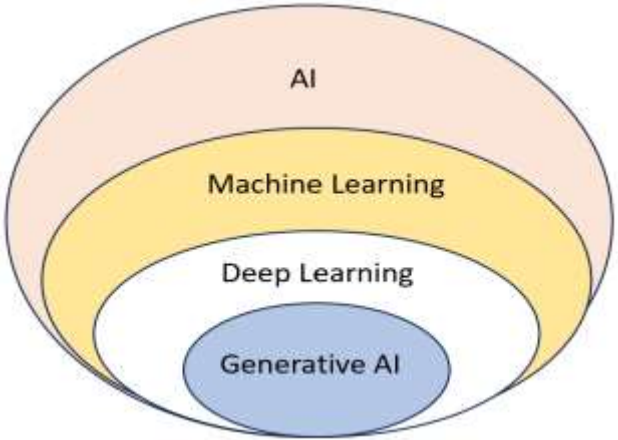
Table of contents

1

What is machine learning?

Machine learning :

- Machine learning is a subset of artificial intelligence that focuses on enabling machines to learn and improve from experience without being explicitly programmed. It involves creating models that can:
- Analyze data.
- Identify patterns.
- Make predictions or decisions based on data.



Machine learning :

- Instead of giving the computer step-by-step instructions, you provide it with data, and it figures out patterns or rules on its own.
- Machine learning is when computers learn from examples, just like how you learn from experience."
- Example: "Imagine teaching a child to recognize animals by showing them pictures and telling them the names. That’s how machines learn too!"
- Imagine teaching a child to recognize animals:

Machine learning :

- You show them pictures of cats and dogs and say, “This is a cat,” or “This is a dog.”

After seeing many examples, the child learns to recognize cats and dogs on their own.

Machine learning works similarly:

- You give the computer lots of data (e.g., pictures of cats and dogs with labels).
- The computer learns patterns from the data (e.g., cats have pointy ears, dogs have floppy ears).

2

Why learning?

Why learning ?

- Unlike traditional programming, where a machine executes predetermined rules written in a program for a specific purpose, machine learning allows machines to independently learn how to interpret contexts and make decisions.
- Machine learning is useful in cases where:
  Human expertise does not exist in certain domains.
  Humans cannot explain how they perform specific tasks (e.g., automatic speech recognition).
- A large volume of data needs to be processed.
  The solution evolves over time (e.g., internet routing).
  The solution must adapt to its user (e.g., biometrics, email filtering).

3

Types of machine learning

Types of machine learning

- Machine learning can be broadly categorized into:
- Supervised Learning: Models are trained on labeled data, learning to map inputs to outputs.
- Unsupervised Learning: Models identify patterns or structures in unlabeled data.

4

Supervised learning

Supervised learning :

- Supervised learning is a type of machine learning where the computer learns from labeled data. Think of it like teaching a child with flashcards:
- You show the child a picture (the data) and tell them what it is (the label).
- After seeing many examples, the child learns to recognize new pictures on their own.
- The computer is given input data (e.g., pictures, numbers, text) and the correct output labels (the answers)
-

Supervised learning :

- It learns to map the input data to the correct output labels by finding patterns or rules.
- Once trained, the computer can predict the correct label for new, unseen data.
- Examples:
- Predicting house prices (regression).
- Classifying emails as spam or not spam (classification).
- Key Idea: The model learns to map inputs to outputs based on labeled examples.

How Does Supervised Learning Work?

- Input Data: You provide the computer with labeled data (e.g., pictures of cats and dogs with labels like "Cat" or "Dog").
- Training: The computer analyzes the data to find patterns (e.g., cats have pointy ears, dogs have floppy ears).
- Prediction: Once trained, the computer can look at new, unlabeled data (e.g., a new picture of an animal) and predict the correct label (e.g., "This is a cat").

5

Unsupervised learning

Unsupervised learning

- Unsupervised learning is a type of machine learning where the computer learns from unlabeled data (data without any answers or labels). Think of it like giving a child a box of mixed toys and asking them to sort the toys into groups without telling them how. The child figures out patterns or similarities on their own.
- The computer is given input data (e.g., pictures, numbers, text) but no labels or answers.
- It analyzes the data to find hidden patterns, structures, or groups.

How Does Unsupervised Learning Work?

- How Does Unsupervised Learning Work?
- Input Data: You provide the computer with unlabeled data (e.g., customer purchase history, pictures of objects).
- Finding Patterns: The computer analyzes the data to find similarities, differences, or groupings.
- Output: The computer organizes the data into clusters, reduces its complexity, or finds hidden structures.

How Does Unsupervised Learning Work?

- Example:
- Image Clustering:
- Input: Pictures of objects without labels.
- Output: Groups of similar images (e.g., cats in one group, dogs in another)

6

Differences between supervised and unsupervised learning

Supervised vs. Unsupervised learning

Aspect	Supervised Learning	Unsupervised Learning
Data	labeled	unlabeled
Goal	predict outputs or classify data	Find patterns or group data
Example	predicting prices, classifying images	clustering custumers, reducing data size
Training process	learns from labeled examples	explores data to find hidden structures
Evaluation	compares predictions to true labels	subjective, based on pattern quality

2

Basic concepts of machine learning

Concepts of machine learning

- Data:

Data is the foundation of machine learning. It can be structured (tables) or unstructured (text, images).

- Types of data: numerical, categorical, textual, images, etc.
- Features:

Features are the input variables used to make predictions.

- Example: to predict house prices, features might include square footage, number of bedrooms, etc.

Concepts of machine learning

- Labels:

In supervised learning, labels are the outputs the model must predict.

- Example: in an image classification dataset, labels are the names of objects (e.g., cat, dog).
- Overfitting:

The model learns the training data too well, including noise, which harms its performance on new data.

Concepts of machine learning

- Solutions: regularization, cross-validation, data augmentation.
- Underfitting:

The model is too simple and fails to capture the underlying patterns in the dataset

- Solutions: increase model complexity, add more features.

Machine learning workflow

1. Data Collection:

Identify data sources (databases, APIs, sensors, etc.)

Ensure the data is representative of the problem to be solved.

2. Data Preprocessing:

- Cleaning: handling missing values, removing outliers.
- Transformation: normalization, encoding categorical variables.
- Dimensionality reduction: PCA, feature selection.

Machine learning workflow

3. Modeling:

- Choose an algorithm based on the problem (classification, regression, clustering).
- Train the model on the training data.
- Tune hyperparameters to optimize performance.

4. Evaluation:

- Use a test dataset to evaluate the model's performance.

Machine learning workflow

Evaluation metrics: accuracy, precision, recall, F1-score (classification); MSE, RMSE (regression).

5. Deployment:

- Integrate the model into an application or system.
- Monitor performance in real-time.
- Update the model with new data if necessary.

Conclusion

- In this chapter we explained essential ideas such as data, features, labels, overfitting, and underfitting
- The Machine Learning Workflow is a step-by-step process, including data collection, preprocessing, modeling, evaluation, and deployment.