

Unsupervised Learning

🎓 Topic: Supervised Learning vs Unsupervised Learning

🔑 Core Concepts

Aspect	Supervised Learning	Unsupervised Learning
Data Labels	Requires labeled data (input → output)	No labels; only raw input data
Goal	Learn mapping from input to output	Discover hidden structure or patterns
Examples	Classification, Regression	Clustering, Dimensionality Reduction
Key Algorithms	Linear Regression, Decision Trees, SVM	K-Means, DBSCAN, PCA, Autoencoders

🧠 Intuition

Supervised Learning

Imagine teaching a child using flashcards:

- One side: picture of an animal 🐶
- Other side: label ("Dog")

After enough examples, they can guess correctly for unseen cards.

Unsupervised Learning

Now imagine handing the child a pile of animal photos with no labels. You ask them to **group similar ones**. They'll make piles of "things with fur", "things with feathers", etc. — discovering structure **without labels**.

📐 Mathematical View

Supervised Learning

You want to approximate a function:

$$f: X \rightarrow Y$$

Where:

- X : Feature space (e.g., image pixels, customer features)
- Y : Target (e.g., "dog", price, yes/no)

Objective:

- Minimize a **loss function**, e.g., Mean Squared Error for regression or Cross-Entropy for classification.

Unsupervised Learning

No Y . The goal is often to **optimize an internal structure**, like:

- **Clustering**: Minimize distance within clusters and maximize between clusters.
- **Dimensionality reduction**: Maximize variance explained (PCA), or minimize reconstruction loss (Autoencoders).

Real-World Examples

Application	Supervised Learning	Unsupervised Learning
Email Filtering	Spam vs Not Spam	Grouping emails by topic
Finance	Predicting credit default (0/1)	Segmenting customers based on behavior
Medical	Diagnosing based on symptoms (disease label)	Grouping patients into risk categories
E-commerce	Predict next product to buy	Finding user clusters for recommendations

Common Misconceptions

- **"Unsupervised learning is worse than supervised"** – Not true. It's different. For unknown labels or exploratory analysis, it's **crucial**.

- **"Unsupervised learning gives you labels"** – No, it groups data. You may need to interpret/label those clusters later.
 - **"Supervised needs a lot of data"** – Yes, especially deep learning. But not all supervised models need massive data.
-

Discussion Questions

1. In your current domain, what problems could be modeled as supervised or unsupervised?
-

K-Means