Machine Learning

Week 1 • Class 5

Types of Learning Problems

Clustering & Optimization

Class Objective

By the end of this class, students will be able to:

- Understand unsupervised learning concepts
- Explain clustering and its use cases
- Understand optimization in machine learning
- Identify real-world examples of both

Types of Learning Problems (Recap)

So far, we have covered:

- Classification
- Regression

This class introduces **unsupervised learning** and **optimization**.

What Is Unsupervised Learning?

In **unsupervised learning**:

- Data is **unlabeled**
- The system discovers patterns automatically
- No predefined output is provided

Clustering is one of the most common unsupervised learning tasks.

What Is Clustering?

Clustering is a learning problem where:

- Data points are grouped based on similarity
- Similar items are placed in the same group (cluster)
- No labels are given beforehand

The goal is to uncover hidden structure in data.

Key Characteristics of Clustering

Clustering problems typically involve:

- Unlabeled data
- Similarity or distance measures
- Grouping based on patterns

Clusters are formed based on feature similarity.

Examples of Clustering Problems

Domain	Clustering Use Case
Marketing	Customer segmentation
Education	Student behavior grouping
Healthcare	Patient risk profiling
Finance	Spending pattern analysis
Image processing	Image segmentation

Common Clustering Algorithms

Popular clustering algorithms include:

- K-Means
- Hierarchical Clustering
- DBSCAN
- Mean Shift

Each algorithm defines similarity differently.

K-Means Clustering

K-Means is a centroid-based clustering algorithm.

Key ideas:

- Each cluster has a center (centroid)
- Data points belong to the nearest centroid
- Centroids are updated iteratively

Characteristics of K-Means

- Fast and efficient for large datasets
- Works well with spherical clusters
- Sensitive to:
  - Initial centroid selection
  - Outliers
  - Feature scaling

Requires the value of **K** beforehand.

Hierarchical Clustering

Hierarchical clustering builds clusters in a tree-like structure.

Produces a **dendrogram** to visualize clustering.

Characteristics of Hierarchical Clustering

- Easy to visualize and interpret
- Does not require K upfront
- Computationally expensive
- Not suitable for very large datasets

Good for exploratory data analysis.

DBSCAN (Density-Based Clustering)

DBSCAN groups data based on density.

Key concepts:

- Dense regions form clusters
- Sparse regions are considered noise
- Clusters can have arbitrary shapes

Characteristics of DBSCAN

- Automatically detects number of clusters
- Handles noise and outliers well
- Works with irregular cluster shapes
- Struggles with varying densities

Widely used in spatial and anomaly detection tasks.

Mean Shift Clustering

Mean Shift is a centroid-based but non-parametric algorithm.

It works by:

- Placing a window around data points
- Shifting the window toward dense regions
- Converging to cluster centers automatically

Characteristics of Mean Shift

- No need to specify number of clusters
- Can find arbitrarily shaped clusters
- Computationally expensive
- Sensitive to bandwidth selection

Often used in image segmentation.

Summary of Clustering Algorithms

Algorithm	Needs K?	Handles Noise	Shape
K-Means	Yes	No	Spherical
Hierarchical	No	No	Tree-based
DBSCAN	No	Yes	Arbitrary
Mean Shift	No	Limited	Arbitrary

Challenges in Clustering

Common challenges include:

- Choosing the right number of clusters
- Handling noise and outliers
- Scaling features
- Interpreting cluster meaning

Clustering results often require domain knowledge.

What Is Optimization?

Optimization in machine learning refers to:

- Finding the best model parameters
- Minimizing error or loss
- Improving model performance

Optimization drives learning in most ML algorithms.

Optimization in Simple Terms

Optimization answers the question:

" How do we make the model as accurate as possible? "

This is done by:

- Defining a loss function
- Minimizing that loss

Loss Function (Concept Overview)

A **loss function** measures:

- How wrong a model's prediction is
- Difference between predicted and actual values

Lower loss means better performance.

Optimization Example

Example: House price prediction

- Model predicts price
- Compare prediction with actual price
- Calculate error
- Adjust model parameters
- Repeat until error is minimized

This process is optimization.

Common Optimization Techniques

Some commonly used techniques include:

- Gradient Descent
- Stochastic Gradient Descent (SGD)
- Mini-batch Gradient Descent

These methods update parameters iteratively.

Optimization vs Clustering

Clustering	Optimization
Groups similar data	Improves model performance
Unsupervised task	Used across ML tasks
No labels required	Uses loss functions
Pattern discovery	Error minimization

Both play critical roles in ML systems.

Real-World Importance

Clustering and optimization are used in:

- Recommendation systems
- Customer analytics
- Image and speech recognition
- Training deep learning models

They form the backbone of modern ML pipelines.

Class Summary

- Clustering groups unlabeled data
- It is a key unsupervised learning task
- Optimization improves model accuracy
- Loss functions guide optimization
- Both are essential in real-world ML

Next Class

Supervised Learning – Concepts and Workflow