Machine Learning

Week 1 · Class 3

Types of Learning Problems

Classification

Class Objective

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

- Understand what a classification problem is
- Identify classification tasks in real-world scenarios
- Distinguish classification from regression
- Recognize common classification algorithms

Types of Learning Problems

In supervised machine learning, problems are commonly divided into:

- Classification
- Regression

This class focuses on **classification**.

What Is Classification?

Classification is a supervised learning problem where:

- The output (label) belongs to a **finite set of categories**
- The model predicts a **class** rather than a number

The goal is to assign the correct label to new data.

Key Characteristics of Classification

Classification problems typically involve:

- Labeled training data
- Discrete output values
- Decision boundaries between classes

Examples include binary and multi-class problems.

Binary Classification

Binary classification involves **two possible classes**.

Examples:

- Spam or Not Spam
- Fraud or Not Fraud
- Pass or Fail
- Disease or No Disease

This is the simplest form of classification.

Multi-Class Classification

Multi-class classification involves **more than two classes**.

Examples:

- Handwritten digit recognition (0–9)
- News article categorization
- Student grade classification (A, B, C, D, F)

Each data point belongs to exactly one class.

Examples of Classification Problems

Problem Domain	Classification Task
Email systems	Spam detection
Finance	Loan approval
Healthcare	Disease diagnosis
Education	Student performance category
Security	Face recognition

Classification vs Regression

Classification	Regression
Predicts categories	Predicts continuous values
Discrete output	Numerical output
Example: Spam/Not Spam	Example: House price

Understanding this distinction is critical.

Common Classification Algorithms

Some widely used classification algorithms:

- Logistic Regression
- k-Nearest Neighbors (k-NN)
- Decision Trees
- Support Vector Machines (SVM)
- Naive Bayes
- Neural Networks

Algorithm choice depends on the problem and data.

Logistic Regression

What It Is

- A simple and widely used classification algorithm
- Used for predicting **two possible outcomes**

How It Works

- Combines input features using weights
- Outputs a probability between 0 and 1
- Uses a threshold to decide the class

When to Use

- When the outcome has **two classes**
- Predicting **Pass / Fail**
- Email **Spam / Not Spam**

k-Nearest Neighbors (k-NN)

What It Is

- A classification algorithm based on similarity
- Does not build a model in advance

How It Works

- Looks at the **k closest data points**
- Assigns the most common class among them

When to Use

- When the dataset is **small**
- When similar data points usually belong to the same class
- Classifying students based on **marks and attendance**
- Recommending items based on **similar users**

Decision Trees

What It Is

- A tree-shaped model that uses decision rules
- Very easy to understand and explain

How It Works

- Splits data using questions
- Each split reduces uncertainty
- Final decision is made at leaf nodes

When to Use

- When decisions must be **clearly explained**
- When rules matter more than accuracy
- **Loan approval** based on income and credit score
- Student eligibility decisions

Support Vector Machines (SVM)

What It Is

- A powerful classification algorithm
- Works well with complex data

How It Works

- Finds the best boundary separating classes
- Maximizes the distance between classes

When to Use

- When data has **many features**
- When classes are clearly separable
- **Face recognition**
- Handwritten character recognition

Naive Bayes

What It Is

- A probability-based classification algorithm
- Assumes features are independent

How It Works

- Calculates probability for each class
- Selects the class with the highest probability

When to Use

- When dealing with **text data**
- When fast training is needed
- **Email spam filtering**
- Sentiment analysis of reviews

Neural Networks

What It Is

- A flexible and powerful learning model
- Inspired by how the human brain works

How It Works

- Data flows through multiple layers
- Learns complex patterns automatically

When to Use

- When the problem is **complex**
- When large amounts of data are available
- Image recognition
- Voice assistants and chatbots

Algorithm Selection (Beginner View)

Problem Type	Suggested Algorithm
Simple binary decision	Logistic Regression
Small, similarity-based data	k-NN
Rule-based decisions	Decision Trees
High-dimensional data	SVM
Text classification	Naive Bayes
Images or speech	Neural Networks

Key Takeaway

- Start with **simple models**
- Choose algorithms based on:
  - Data size
  - Problem complexity
  - Need for explanation

More advanced models will be explored later.

Classification Output

Classification models may output:

- A predicted class label
- Probabilities for each class

Example:

- Spam (0.92)
- Not Spam (0.08)

Decision Boundary (Conceptual)

A **decision boundary** separates different classes.

- Learned during training
- Depends on features and model
- Can be linear or non-linear

We will visualize decision boundaries in later labs.

Evaluating Classification Models

Common evaluation metrics include:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

Proper evaluation ensures model reliability.

Classification Workflow

1. Define the classification problem
2. Collect and label data
3. Select features
4. Split data (training/testing)
5. Train classification model
6. Evaluate performance

This mirrors the general ML pipeline.

Real-World Considerations

When solving classification problems:

- Data may be imbalanced
- Misclassification costs vary
- Accuracy alone may be misleading

These issues will be addressed in future classes.

Class Summary

- Classification predicts categorical outputs
- Can be binary or multi-class
- Uses labeled data
- Requires appropriate evaluation metrics
- Widely used in real-world applications

Next Class

Types of Learning Problems – Regression