Machine Learning

Week 1 · Class 4

Types of Learning Problems

Regression

Class Objective

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

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

Types of Learning Problems

In supervised machine learning, common problem types include:

- Classification
- Regression

This class focuses on **regression**.

What Is Regression?

Regression is a supervised learning problem where:

- The output (label) is a **continuous numerical value**
- The model predicts quantities rather than categories

The goal is to estimate an accurate numeric outcome.

Key Characteristics of Regression

Regression problems typically involve:

- Labeled training data
- Continuous output values
- Error measured as distance from true value

Examples include price, score, or demand prediction.

Examples of Regression Problems

Problem Domain	Regression Task
Real estate	House price prediction
Finance	Stock price forecasting
Education	Student score prediction
Healthcare	Blood pressure estimation
Weather	Temperature prediction

Regression vs Classification

Regression	Classification
Predicts numerical values	Predicts categories
Continuous output	Discrete output
Example: House price	Example: Spam detection

Understanding this distinction is essential.

Simple vs Multiple Regression

Simple Regression

- Uses one input feature
- Example: Price vs area

Multiple Regression

- Uses multiple input features
- Example: Price vs area, location, rooms

Linear Regression (Concept Overview)

Linear regression models:

- A linear relationship between inputs and output
- A straight-line fit to the data

It is one of the most widely used regression techniques.

Non-Linear Regression (Concept Overview)

Non-linear regression models:

- Complex relationships
- Curved or irregular patterns
- Real-world data behavior

Often handled using advanced models.

Common Regression Algorithms

Popular regression algorithms include:

- Linear Regression
- Polynomial Regression
- Ridge and Lasso Regression
- Decision Tree Regression
- Random Forest Regression
- Neural Networks

Linear Regression

What It Is

- The simplest and most commonly used regression algorithm
- Models a straight-line relationship between inputs and output

How It Works

- Fits a straight line that best represents the data
- Minimizes the difference between predicted and actual values

When to Use

- When the relationship is roughly linear
- When you want a simple and interpretable model
- Predicting **house price based on area**
- Estimating **salary based on years of experience**

Polynomial Regression

What It Is

- An extension of linear regression
- Can model curved relationships

How It Works

- Transforms input features into polynomial terms
- Fits a curved line instead of a straight line

When to Use

- When data shows a **non-linear trend**
- When linear regression underfits
- Predicting **growth trends over time**
- Modeling **exam score vs study hours** where improvement slows

Ridge and Lasso Regression

What They Are

- Regularized versions of linear regression
- Help prevent overfitting

How They Work

- Penalize large coefficients
- Reduce model complexity

When to Use

- When there are **many input features**
- When features are correlated
- Predicting **house price using many property features**
- Financial prediction models with many variables

Decision Tree Regression

What It Is

- A tree-based regression model
- Uses rules to make predictions

How It Works

- Splits data based on feature conditions
- Outputs average value at leaf nodes

When to Use

- When relationships are **non-linear**
- When explanation is important
- Predicting **insurance cost** based on age and risk factors
- Estimating **sales based on season and region**

Random Forest Regression

What It Is

- An ensemble of multiple decision trees
- Produces more stable predictions

How It Works

- Builds many trees on random subsets of data
- Averages predictions from all trees

When to Use

- When accuracy is more important than simplicity
- When data is noisy
- Predicting **house prices with many influencing factors**
- Demand forecasting for products

Neural Networks (Regression)

What It Is

- A powerful and flexible regression model
- Can learn very complex patterns

How It Works

- Passes data through multiple layers of neurons
- Adjusts weights to minimize prediction error

When to Use

- When data is **large and complex**
- When simpler models fail
- Predicting **energy consumption**
- Stock price trend estimation

Regression Algorithm Comparison (Beginner View)

Problem Type	Suggested Algorithm
Simple linear relationship	Linear Regression
Curved trend	Polynomial Regression
Many correlated features	Ridge / Lasso
Rule-based numeric prediction	Decision Tree Regression
High accuracy needed	Random Forest
Complex patterns, large data	Neural Networks

Key Takeaway

- Start with **simple regression models**
- Increase complexity only when needed
- Algorithm choice depends on:
  - Data size
  - Feature complexity
  - Interpretability needs

Understanding regression algorithms builds a strong ML foundation.

Regression Output

Regression models produce:

- A single numerical prediction
- Continuous-valued output

Example:

- Predicted house price: 7,500,000

Evaluating Regression Models

Common evaluation metrics include:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R-squared (R^2)

Lower error indicates better performance.

Regression Workflow

1. Define the regression problem
2. Collect and prepare data
3. Select relevant features
4. Split data (training/testing)
5. Train regression model
6. Evaluate predictions

This workflow mirrors general ML practice.

Real-World Considerations

When working with regression problems:

- Outliers can distort results
- Data scaling may be required
- Linear assumptions may not hold

These will be addressed in later sessions.

Class Summary

- Regression predicts continuous values
- Uses labeled data
- Can be linear or non-linear
- Requires numerical evaluation metrics
- Widely used in prediction tasks

Next Class

Types of Learning Problems – Clustering & Optimization