Linear Regression Problem Definition

Goal

Model relationship between input X and output Y

Core Assumption

Linear relationship

$$Y = f(X) + \varepsilon$$

Training Data

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

n labeled examples

Objective

Find function f that best approximates the true relationship

Learning Paradigm

Supervised Learning:

Learning from labeled examples

Regression Workflow

1. Collect Data

Training examples (X, Y)

↓

2. Learn Function f

Find best fit to data

↓

3. Predict

Given new x → estimate $\hat{y} = f(x)$

Applications

- Price prediction
- Trend forecasting
- Causal inference

Bridge from theory to practical machine learning

Matrix Representation

Linear Model

For multiple features:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon$$

In matrix form:

$$Y = X\beta + \varepsilon$$

where:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Y: $n\times1$, X: $n\times(p+1)$, β: $(p+1)\times1$, ε: $n\times1$

Least Squares Solution

Objective: Minimize sum of squared errors

$$\min ||Y - X\beta||^2$$

Normal Equation:

$$X^T X\beta = X^T Y$$

Solution:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Prediction: $\hat{Y} = X\hat{\beta}$

This requires $X^T X$ to be invertible
(X has full column rank)