Linear Regression

Overview

Linear regression is one of the simplest and most widely used algorithms in machine learning.
It predicts a continuous outcome (like price or temperature) by finding the best-fitting straight line through data points.

In essence, it asks:

> "How does one variable (X) influence another (Y)?"

It is an excellent starting point for understanding how machines learn patterns from numerical data.

Key Concepts

- **Dependent Variable (Y):** The value we aim to predict (e.g. house price).
- **Independent Variable (X):** The input used to make the prediction (e.g. size in square metres).
- **Line of Best Fit:** The line that minimises the total difference between predicted and actual values.
- **Loss Function:** Measures prediction error (commonly *Mean Squared Error*).

Visual Summary

Data Points → Best Fit Line → Prediction

Think of a scatter plot where each dot represents a house.
A single line passes through them, showing how price increases with size.

Real-World Applications

- 🏡 **Real Estate:** Predicting house prices based on size or location.
- 🚗 **Ride Pricing:** Estimating Uber surge prices.
- 💰 **Finance:** Predicting market returns or sales trends.
- ☀️ **Weather Forecasting:** Estimating temperature or rainfall amounts.

Example

Price = 25,000 + (300 × House Size) If the model learns that each additional square metre increases the price by £300,
it can predict the cost of future houses by applying that formula.

Why It Matters

Linear regression forms the foundation of more complex AI models.

Even advanced systems often rely on regression principles to estimate relationships between variables.