ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DAY – 13

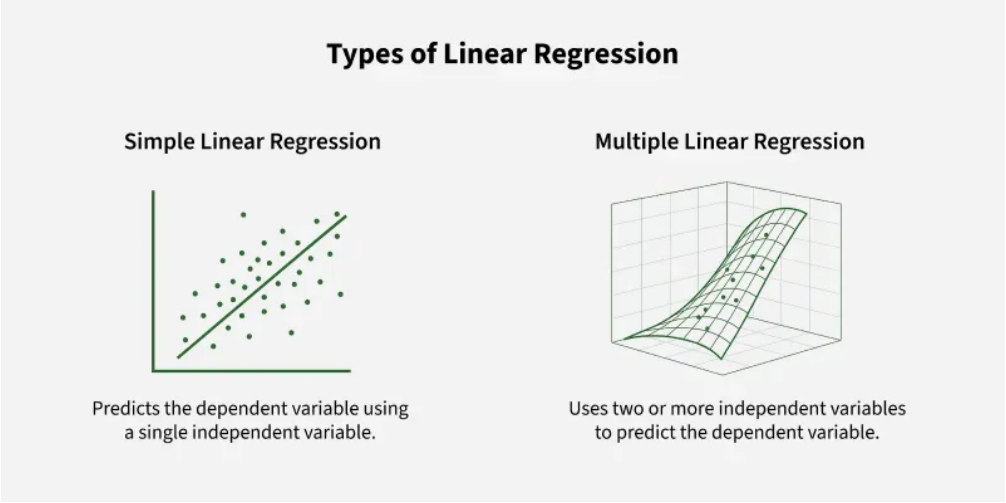
Date: Jul 09, 2025

Linear Regression

Linear regression is a type of supervised machine-learning algorithm that learns from the labelled datasets and maps the data points with most optimized linear functions which can be used for prediction on new datasets. It assumes that there is a linear relationship between the input and output, meaning the output changes at a constant rate as the input changes. This relationship is represented by a straight line.

For example we want to predict a student's exam score based on how many hours they studied. We observe that as students study more hours, their scores go up. In the example of predicting exam scores based on hours studied. Here

- **Independent variable (input):** Hours studied because it's the factor we control or observe.
- **Dependent variable (output):** Exam score because it depends on how many hours were studied.



Types of Linear Regression

When there is only one independent feature it is known as Simple Linear Regression or Univariate Linear Regression and when there are more than one feature it is known as Multiple Linear Regression or Multivariate Regression.

1. Simple Linear Regression

Simple linear regression is used when we want to predict a target value (dependent variable) using only one input feature (independent variable). It assumes a straight-line relationship between the two.

Formula:

$y^=\theta_0+\theta_1x$

Where: $y^$ is the predicted value, x is the input (independent variable), θ_0 is the intercept (value of $y^$ when x=0), θ_1 is the slope or coefficient (how much $y^$ changes with one unit of x)
Example: Predicting a person's salary (y) based on their years of experience (x).

2. Multiple Linear Regression

Multiple linear regression involves more than one independent variable and one dependent variable. The equation for multiple linear regression is:

$y^=\theta_0+\theta_1x_1+\theta_2x_2+\cdots+\theta_nx_n$

where: $y^$ is the predicted value, $x_1,x_2,\ldots,x_nx_1$,x_2 ,…,x_n are the independent variables
$\theta_1,\theta_2,\ldots,\theta_n\theta_1$,θ_2 ,…,θ_n are the coefficients (weights) corresponding to each predictor.
θ_0 is the intercept.
The goal of the algorithm is to find the best Fit Line equation that can predict the values based on the independent variables.

Use Case of Multiple Linear Regression

Multiple linear regression allows us to analyze relationship between multiple independent variables and a single dependent variable. Here are some use cases:

- **Real Estate Pricing:** In real estate MLR is used to predict property prices based on multiple factors such as location, size, number of bedrooms, etc. This helps buyers and sellers understand market trends and set competitive prices.
- **Financial Forecasting:** Financial analysts use MLR to predict stock prices or economic indicators based on multiple influencing factors such as interest rates, inflation rates and market trends. This enables better investment strategies and risk management24.
- **Agricultural Yield Prediction:** Farmers can use MLR to estimate crop yields based on several variables like rainfall, temperature, soil quality and fertilizer usage. This information helps in planning agricultural practices for optimal productivity
- **E-commerce Sales Analysis:** An e-commerce company can utilize MLR to assess how various factors such as product price, marketing promotions and seasonal trends impact sales.