

Intro to Data Science Assignment #03

Name: Muhammad Nadeem

Roll: 201980050

Comparison of Linear and Multi-Regression on the basis of why, what, when, How?

Linear Regression:

What is it?

Linear regression is a statistical method used to model the relationship between a dependent variable (target) and one or more independent variables (predictors) by fitting a linear equation to the observed data. The objective is to find the best-fitting straight line that minimizes the distance between the predicted values and the actual values.

Why is it used?

Linear regression is commonly used for

several reasons:

1. Prediction:

It can be used to predict the values of the dependent variable when the values of independent variables are known.

2. Relationship Assessment:

It helps to determine the strength and direction of the relationship between the variables.

3. Understanding Causality:

It aids in understanding how changes in independent variables can impact the dependent variable.

4. Identifying Important Variables: It allows us to identify which independent variables have a significant impact on the dependent variable.

How is it done?

In simple linear regression (one independent variable), the equation takes the form:

$$y = mx + b$$

where:

- y is the dependent variable (target)
- x is the independent variable (predictor)
- m is the slope of the line (representing the relationship between x and y)
- b is the y -intercept (the value of y when x is 0)

The best-fitting line is determined by minimizing the sum of squared differences between the actual values and the predicted values (least squares method).

When is it used?

Linear regression is used when there is a linear relationship between the dependent and independent variables. It is applicable when the data points approximately form a straight line on a scatter plot.

Multilinear Regression:

What is it?

Multilinear regression, also known as **multiple linear regression**, is an version of linear regression that involves multiple independent variables to predict the dependent variable. Instead of a straight line, it fits a hyperplane in a higher-dimensional space.

Why is it used?

Multilinear regression is used for similar reasons as simple linear regression, but it allows for more complex models that consider multiple factors that can influence the dependent variable simultaneously.

How is it done?

The equation for multilinear regression is:

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

where:

- y is the dependent variable (target)

- x_1, x_2, \dots, x_n are the independent variables (predictors)
- $b_0, b_1, b_2, \dots, b_n$ are the coefficients (slopes) for each independent variable

The goal is to find the **best-fitting hyperplane** by minimizing the sum of squared differences between the **actual values** and the **predicted values**.

When is it used?

Multilinear regression is used when there are multiple independent variables that may jointly influence the dependent variable. It is suitable when the relationship between the dependent and independent variables is not a simple straight line but can be approximated by a hyperplane.

In both linear and multilinear regression, it is important to check for assumptions such as linearity, independence of errors, constant variance, and normality of