1.What is linear regression in machine learning?

Linear regression is a type of machine learning algorithm used to model the relationship between a dependent variable (target) and one or more independent variables (predictors). The goal of linear regression is to find the line of best fit that represents the relationship between the variables.

2.What are linear regression assumptions?

Linear regression is a statistical model that relies on a few assumptions about the data. These assumptions are important to ensure that the results of the model are accurate and reliable.

They are five types of assumptions:

- **Linearity:** The relationship between the dependent variable and the independent variable(s) is linear.
- **Independence:** The observations in the data set are independent of each other. This means that the value of one observation does not depend on the value of another observation.
- **Homoscedasticity:** The variance of the errors is constant across all levels of the independent variable(s).
- **Normality:** The errors are normally distributed. This means that the distribution of the residuals is approximately normal.
- **Multicollinearity:** The independent variables are not highly correlated with each other. This means that there is no linear relationship between any pair of independent variables.

I. ASSUMPTION-LINEARTIY

The relationship between the dependent variable and the independent variable(s) should be linear, which means that the slope of the line should remain constant for all values of the independent variable(s). If this assumption is violated, then linear regression may not be the appropriate model to use.

II. ASSUMPTION-INDEPEDENCE

Each observation should be independent of the others. This means that the value of one observation should not be influenced by the value of another observation. If this assumption is violated, then the model's results may not be valid.

III. ASSUMPTION-HOMOSCEDASTICITY

The variance of the errors (also called residuals) should be constant across all values of the independent variable(s). In other words, the spread of the residuals should be the same for all levels of the independent variable(s). If this assumption is violated, then the model may not accurately capture the true relationship between the dependent variable and the independent variable(s).

IV. ASSUMPTION-NORMALITY

The errors (residuals) should be normally distributed. In other words, the distribution of the residuals should be a bell-shaped curve with a mean of zero. If the residuals are not normally distributed, then the model may not be appropriate for the data.

V. ASSUMPTION-MULITYCOLLINEARITY

The independent variables should not be highly correlated with each other. This is called multicollinearity, and it can lead to unstable estimates of the coefficients. If there is multicollinearity, then the model may not accurately capture the true relationship between the dependent variable and the independent variable(s).