**What is linear regression?**

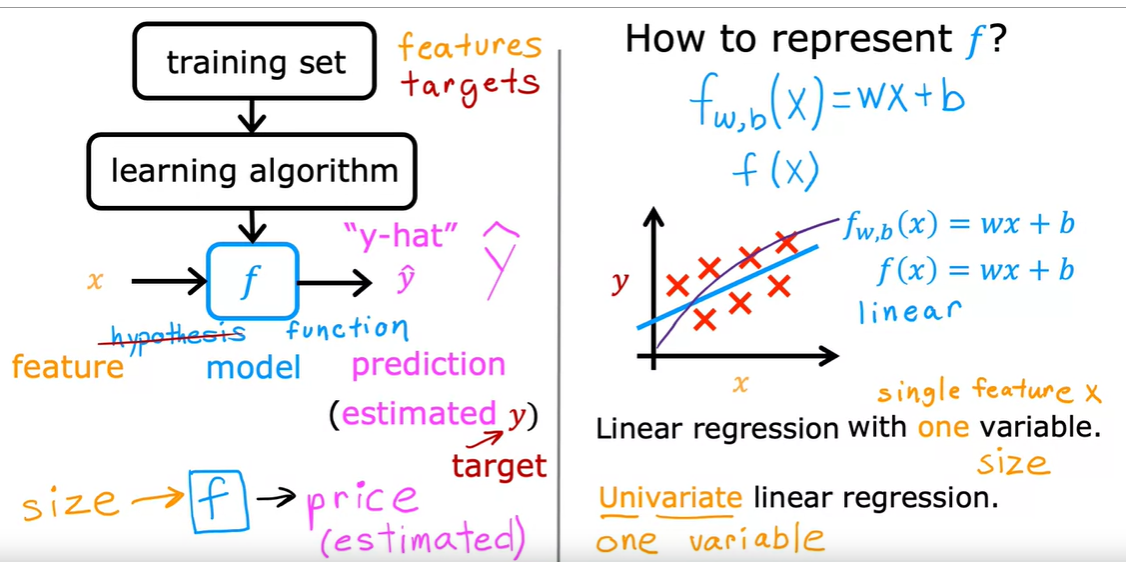
**For linear equation maths:** [**Linear\_quation\_y =mx + c**](https://thirdspacelearning.com/gcse-maths/algebra/y-mx-c/%23:~:text=The%20equation%20y%20%3D%20mx%20%2B%20c%20is%20the%20general%20equation%20of,crosses%20the%20y%20%2Daxis).)

Linear regression is a type of supervised machine learning algorithm used to predict a continuous output variable based on one or more input variables. The goal of linear regression is to find the linear relationship between the input variables (also known as independent variables or features) and the output variable (also known as the dependent variable), so that we can use this relationship to make predictions on new data.

In simple linear regression, we have only one input variable and one output variable, and we are trying to find the straight line that best fits the data points. The equation of this line is given by Y = a + bX, where Y is the output variable, X is the input variable, a is the intercept, and b is the slope.

In multiple linear regression, we have more than one input variable, and the equation becomes Y = a + b1X1 + b2X2 + ... + bnXn. Here, we are trying to find the values of the intercept and slopes that best fit the data points.

To find the best values of a and b, we use a technique called ordinary least squares (OLS) regression. OLS regression minimizes the sum of the squared differences between the predicted values and the actual values of the output variable. Once we have the values of a and b, we can use the regression equation to make predictions on new data.

One variable linear regression:  
 - f(y) = mx + c  


**what if interviewer ask how do you know slope?**

If an interviewer were to ask how we know the slope in linear regression, we use a technique called least squares regression to estimate the values of the coefficients (including the slope) that minimize the sum of the squared errors between the predicted values and the actual values.

In simple linear regression, the slope can be calculated as the covariance between the input variable and the output variable divided by the variance of the input variable. This gives us the change in the output variable for every unit change in the input variable

In multiple linear regression, the slope is the change in the output variable for every unit change in the input variable, holding all other input variables constant. We use matrix algebra to estimate the values of the coefficients in multiple linear regression

**what are the assumption in linear regression model ?**

Linear regression makes several assumptions about the relationship between the input variables and the output variable. Violation of these assumptions can lead to biased or inefficient estimates of the regression coefficients and unreliable predictions. The key assumptions of linear regression include:

Linearity: The relationship between the input variables and the output variable is linear. In other words, the slope of the regression line is constant across all levels of the input variables.

Independence: The observations in the dataset are independent of each other. In other words, there is no relationship between the residuals (the differences between the predicted and actual values) for different observations.

Homoscedasticity: The variance of the residuals is constant across all levels of the input variables. In other words, there is no pattern in the residuals that varies systematically across the range of the input variables

Normality: The residuals are normally distributed. In other words, the distribution of the residuals is symmetric and bell-shaped.

No multicollinearity: In multiple linear regression, there is no high correlation between any pair of input variables. High correlation between input variables can lead to unstable estimates of the regression coefficients.

No outliers: There are no extreme values in the dataset that have a disproportionate influence on the estimates of the regression coefficients.

These assumptions are important to consider when performing linear regression and interpreting its results. In practice, it is often a good idea to check these assumptions using diagnostic plots and statistical tests before making any conclusions based on a linear regression model.