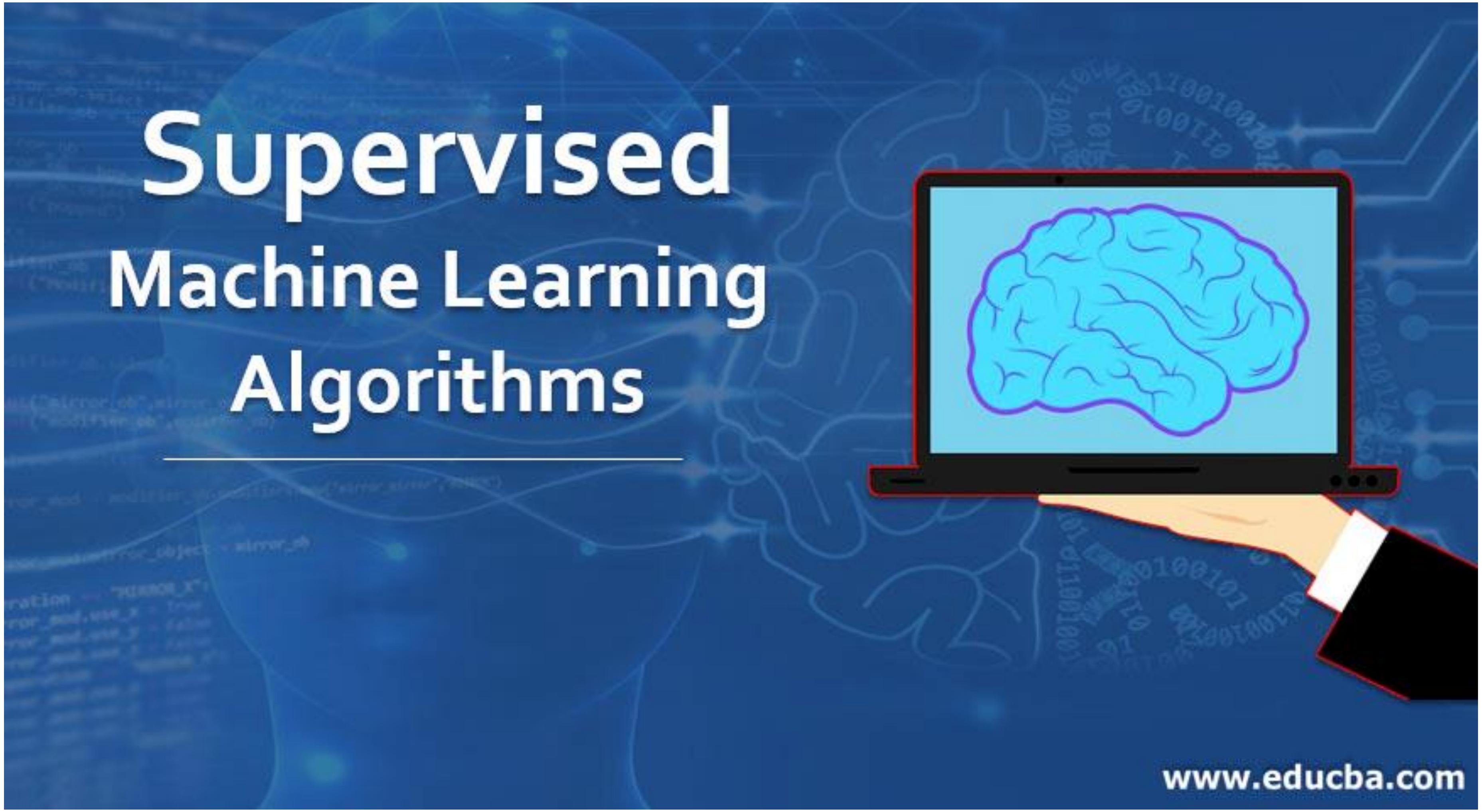

Supervised
Machine Learning
Algorithms
www.educba.com

Last lecture reminder

We learned about:

- The inverse matrix
- Solving linear equations using the inverse matrix
- Mathematical concepts (equation degree, gradient)
- Statistical distributions (Uniform, Binomial, Normal)
- Statistical concepts (Mean, Standard deviation, Variance)

Introduction To Machine Learning

Machine learning (ML) → Machine Learning is a subfield of artificial intelligence that uses algorithms and statistical models to enable computers to make decisions or predictions without being explicitly programmed to perform the task.
It revolves around the idea that machines should be able to learn and adapt through data analysis.
Machine learning models are built by inputting a large amount of data along with the processing of that data to create a model that can make accurate predictions.

The problem we want to solve using machine learning will be with the following structure:

Given <u>features</u> from a dataset obtain a desired <u>label</u>.

In ML features are the different measurements (the dataset points) and label is the forecasting result.

Introduction To Machine Learning

For example → Let’s say we want to forecast what is the fair price to sell a house.

In order to do it we need to give data about the different properties of the house we want to sell (like area, number of bedrooms, bathrooms, address, etc…)

Each of those properties is called a feature and the sell price that will be decided called a label.

In a human world, we as humans will decide what are the more important features and based on that building a model to decide what is the fair price to sell the house.

In ML world the machine should automatically decide which features are more important than others.

Note: When working with machine learning algorithms, it's uncommon to need to construct a new model from scratch. In most cases, we will need to select an existing model that best suits the problem that we are trying to solve.

Introduction To Machine Learning

Machine learning can be categorized into three types:

Supervised Learning → This is a type of machine learning where an algorithm learns from example data and associated output provided during training. The algorithm makes predictions based on the input data, and each input instance comes with the correct output. Over time, the model is able to accurately predict the output when presented with new data. This method is commonly used in applications where historical data predicts likely future events. For instance, it can anticipate when credit card transactions are likely to be fraudulent or when an insurance claim is likely to be valid or invalid.

Unsupervised Learning → Unsupervised learning works differently as it deals with data sets without historical labels. The system is not told the "right answer", instead it automatically finds patterns and relationships in the data. The goal is to explore the underlying structure or data distribution in the dataset.

Introduction To Machine Learning

Unsupervised learning most commonly used for clustering population in different groups which is widely used in market segmentation, or for detecting abnormal behavior, like fraud detection in credit card transactions.

Reinforcement Learning → Reinforcement learning is a kind of machine learning where an agent learns
how to behave in an environment by performing actions and seeing the results. In reinforcement learning,
a software agent makes observations and takes actions within an environment, and in return, it receives rewards. Its objective is to learn to perform actions that maximize some cumulative reward. One of the most common uses of reinforcement learning is in game theory/strategy, it's used to train AI to play different games by rewarding "winning" strategies and penalizing "losing" ones. An example of this was Google's DeepMind AlphaGo program that was trained to play the board game Go.

Introduction To Supervised Learning

When working with supervised learning we have the following data:

1. **Historical date** → Known data from the past.
2. **Labeled data** → The results (labels) from the historical data is also known.

For example → If we want to use ML to decide what is a fair price for selling a given house we can collect historical data of all the houses that has been sold from the past few years (features) and the selling price of each of those houses (label).
Because we have both historical data and the data is labeled we can solve this problem using supervise learning algorithms.

Introduction To Supervised Learning

In supervised learning there are 2 types of labels:

1. **Categorical label** → This type of data is characterized by values that can be divided into several categories but having no order or priority. For instance, consider an example of a survey asking which brand of soft drinks is preferred. The possible answers could be "Brand A", "Brand B", "Brand C" and so forth. These represent different categories, but they do not have any numerical significance and one cannot be considered greater than the other.

2. **Continuous label** → continuous data is information that can be measured on a continuum or a scale and can take on any value within a finite or infinite interval.
   For example, the height of a person, weight of a person, the temperature of a day, the age of a person, etc. These are all continuous data because they can be measured and can represent different values on a particular scale.

Introduction To Supervised Learning

By finding the right label type, we also decide what is the task that we are trying to solve:

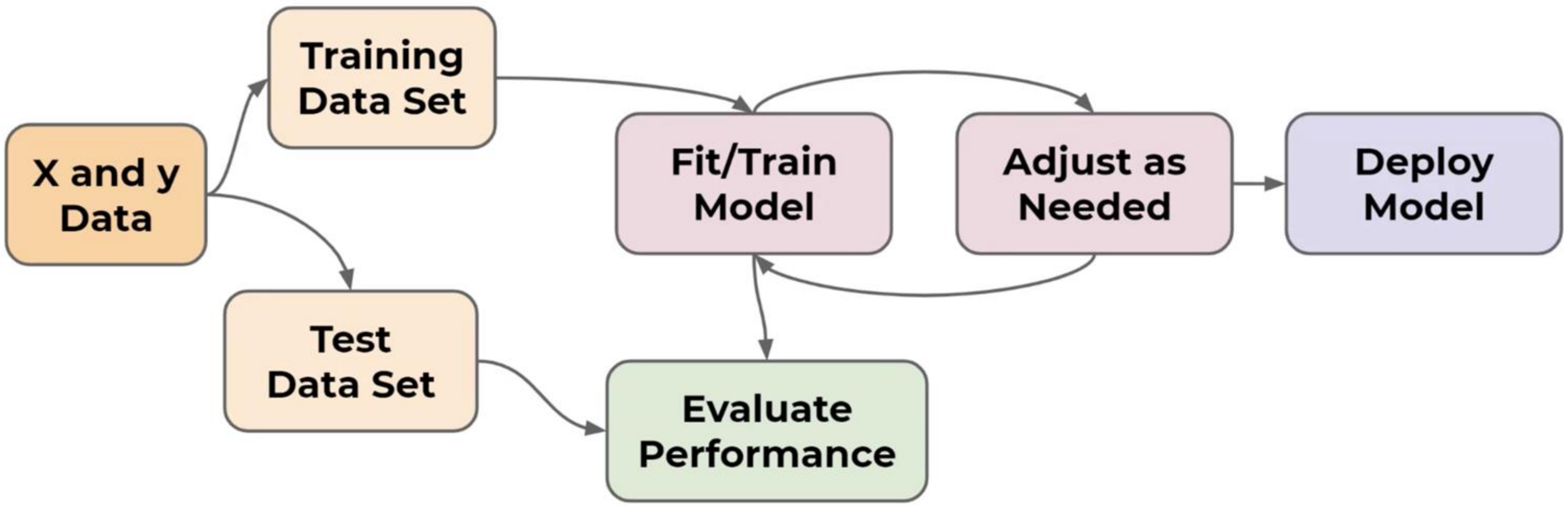
For dataset with categorical label → The task that our ML will solve will be a **classification problem**.

For dataset with continuous label → The task that our ML will solve will be a **regression problem**.

In our selling house example the historical data label is a continuous label type (because the prices are continuous and in theory can have infinite range of prices), so the problem that our ML trying to solve (find the fair price to sell a house) is a regression problem.

Supervised Learning Process In ML

The following flowchart describe the supervised machine learning process:



X → Representing the features

Y → Representing the label

Supervised Learning Process In ML

Flowchart explanation:

In supervised learning we have historical labeled data and we want to use this data to successfully predict future labels with future unseen data.

In order to accomplish that we first split our historical data into 2 parts:

- **Training dataset** → Historical data that our machine learning algorithm will use for the training.
- **Test dataset** → Historical data that will allow us to test our machine learning predictions and see how much they close / not close to the actual results.

We can repeat the training process as much as we want until we will get optimal predictions according to the test dataset (the adjust as needed step).

Once we are happy with our ML prediction results we can deploy the model and let it run on real unseen data (the deployment step).

Introduction To Linear Regression

Linear regression is a type of statistical analysis that attempts to show a relationship between two variables (x & y). Linear regression looks at various data points and plots a line through them, which best expresses this relationship. More specifically, the line is able to predict the dependent variable (y) with the highest level of accuracy, by knowing the independent variable (x).

Linear regression modeling is based on the idea that if a certain type of correlation or relationship (specifically, a linear one) exists between two variables (x and y) in past data, this relationship can be used to predict future data points.

For example → By using a linear regression model, we can analyze the past data. Suppose our analysis reveals that every incremental $100 spent on advertising leads to an increase in sales by $500. This vital insight allows us to strategically plan our advertising budget for the future.

Introduction To Linear Regression

Let's use our house selling example to better understand how linear regression model works.

Given the following historical data on houses selling prices we want to use linear regression to help us predict what should be the fair price of a specific house:

X			y
Area m^2	**Bedrooms**	**Bathrooms**	**Price**
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

ECOM SCHOOL
המכללה למקצועות הדיגיטל וההייטק

Introduction To Linear Regression

In our example the features that we have are:

- Arena of the house (X1)
- Number of bedrooms (X2)
- Number of bathrooms (X3)

And the label is the selling price of each house (Y)

We can also describe this data as the following:

Feature number

Feature number data

Corresponding Label data

X			y
$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$	$\mathbf{y}$
$x^1{}_1$	$x^1{}_1$	$x^1{}_1$	y_1
$x^2{}_1$	$x^2{}_1$	$x^2{}_1$	y_2
$x^3{}_1$	$x^3{}_1$	$x^3{}_1$	y_3
$x^4{}_1$	$x^4{}_1$	$x^4{}_1$	y_4
$x^5{}_1$	$x^5{}_1$	$x^5{}_1$	y_5

Introduction To Linear Regression

Now, we can say that each feature has a different influence on the label value, some has positive influence, some negative influence and some features don’t have any influence at all.

Beta Coefficient (β) → The influence of a specific feature on the label value called the beta coefficient. By finding the beta coefficient of each feature we could get a prediction formula that will help us determine for given new features data what will be the expected label.

In our house prices example we can represent this linear equation as the following:

y	X		
y	x_1	x_2	x_3

$$\hat{y} = \beta_0 x_0 + \cdots + \beta_n x_n$$

ECOM SCHOOL
המכללה למקצועות הדיגיטל וההייטק

Linear Regression Equation

Linear regression equation can be written in the following form:

$$\hat{y} = \beta_0 x_0 + \cdots + \beta_n x_n$$

X → Each X represent a different feature in our dataset (for example area, bedrooms, bathrooms ext…)

β → Each β represent the corresponding coefficient for that feature, provide as a way to understand how much this feature effect the label estimate.

Y-hat (ŷ) → y-hat represents the predicted value of the dependent variable (y), for any given value of the

independent variable (x). It is obtained from the regression line fitted to the data. It is called y-hat because it is an estimate of the true y value, and the hat symbol (^) represents an estimate.

Simple Linear Regression Equation

β0 → This is the y-intercept of the regression line. It is the value of ŷ when x = 0, meaning it's the estimated value of the dependent variable when all independent variables are 0.

In linear regression model with only 1 feature (single X) we are getting the following equation:

Estimated (or predicted) Y value for observation i

Estimate of the regression intercept

Estimate of the regression slope

Value of X for observation i

$$\hat{Y}_i = b_0 + b_1 X_i$$

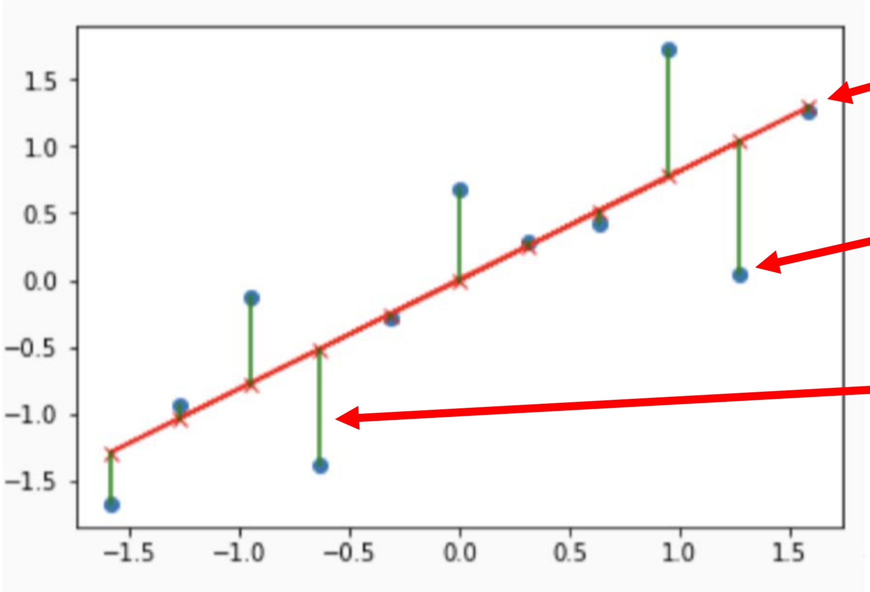
This is very similar to the linear line equation:

$$\boxed{y = mx + b}$$

Simple Linear Regression Equation

In a model with only 1 feature we are getting the equation of the line of best fit, represents a statistical model used to understand the relationship between two variables (x & y).

The aim of linear regression is to find the best-fitting straight line through the data points.
The best-fitting line is the one that minimizes the sum of the squared differences between the actual and the predicted values.



The Linear regression line

Observations (real label values)

The error (the difference between the actual label value and the predicted y-hat value)

Multiple Linear Regression Equation

In a model with multiple features the linear regression equation becomes a multiple linear regression. The general form of the equation is:

$$\hat{y} = \beta_0 x_0 + \cdots + \beta_n x_n$$

This equation can also be written as the following:

$$\hat{y} = \sum_{i=0}^{n} \beta_i x_i$$

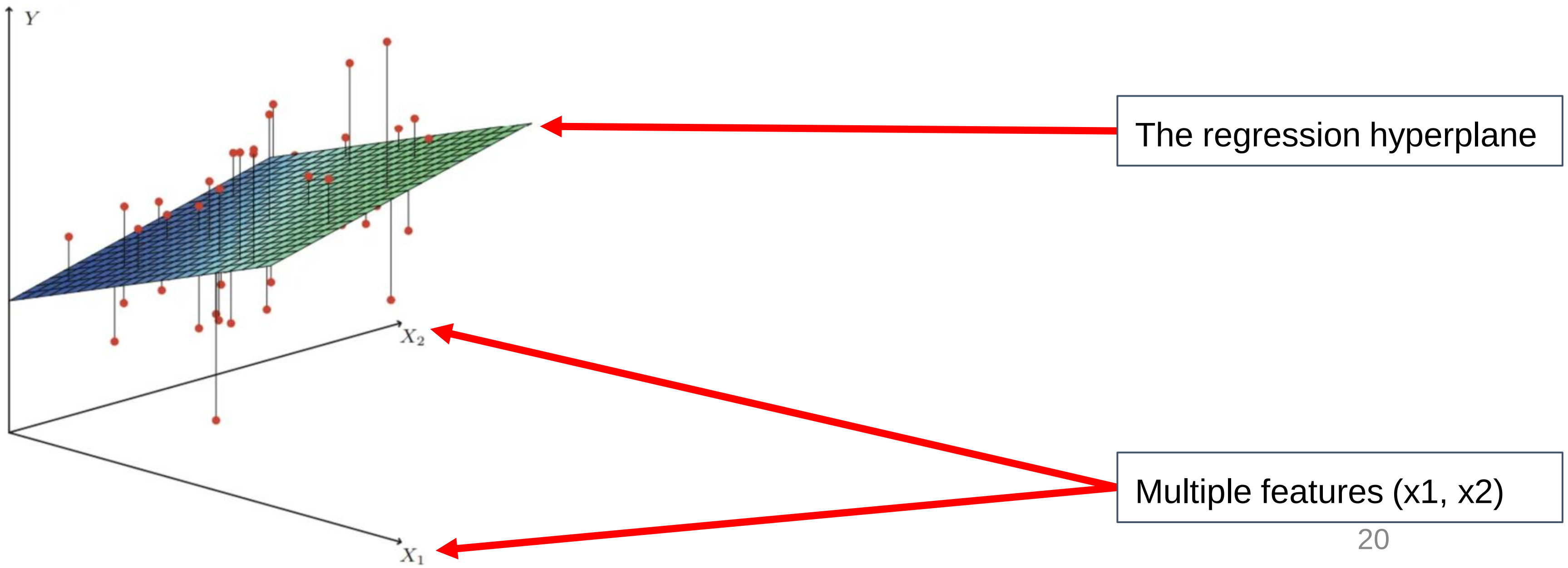
The aim is still to find the optimal values for the β coefficients that minimize the sum of the squared differences between each observation's actual and predicted values. These values are normally computed using methods like the method of least squares.

Multiple Linear Regression Equation

When trying to visualizing a multiple linear regression it become much more difficult than visualizing the simple linear regression line because now we have multiple feature (meaning multiple dimension) and each feature can affect differently on the label value.

So instead of a linear line we are getting what called a **regression hyperplane** that best fits the data.

Each independent variable forms a different axis in this multidimensional space.



When Can We Use Linear Regression

So as we learned, linear regression modeling allow us to use historical data observations in order to predict future label values given one or multiple features.

We can use linear regression modeling only when the following conditions are met:

1. **Linear Relationship** → There should be a linear relationship between the dependent and independent variables. You can often check this condition by creating a scatterplot of each independent variable versus the dependent variable and seeing if a linear pattern appears.
2. **Independence** → The observations should be independent of each other. In other words, the residuals (the differences between the observed and predicted values) at any given point in the dataset should not depend on the residuals at any other point.
3. **Homoscedasticity** → The variance of the errors should be the same across all levels of the independent variables. If the variance changes, then linear regression is not the most appropriate model to use.
4. **Normality** → The residuals of the model should be normally distributed.

When Can We Use Linear Regression

Some of the conditions we can check before performing linear regression modeling like checking linear relationship between the dependent and independent variables.

But some will required finding the residual (the differences between the observed and predicted values) which only can be found after performing linear regression modeling on the data.

So there could be some cases where we will perform linear regression modeling only to find that its not the right modeling for our data.

Note: It's important to remember that the assumptions of linear regression are ideal conditions. In real-world data analysis and predictive modeling, some deviation from these assumptions is often tolerable as long as it’s not too severe and the primary conclusions of the analysis are not threatened. However, severe violations of these assumptions can lead to biased, inefficient, or misleading results.