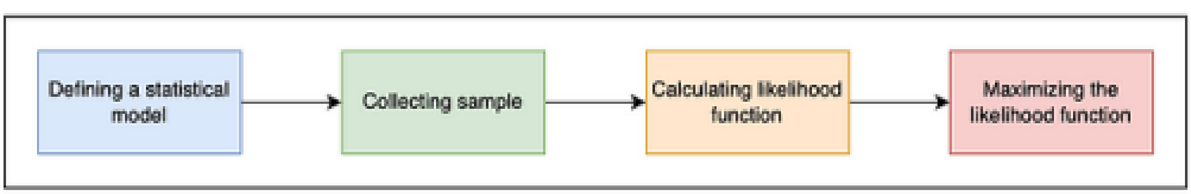
UNIT - 2 Linear models for regression

1.Maximum likelihood estimation (MLE)

Maximum likelihood estimation (MLE) is a statistical approach that determines the models' parameters in machine learning. The idea is to find the values of the model parameters that maximize the likelihood of observed data such that the observed data is most probable.

Let's look at an example to understand MLE better. Assume that we want to estimate the average height of a city's population. However, because of the sheer size of the population, we cannot calculate the true average height of the population. So, we estimate the average height as follows:

- **Defining a statistical model:** We start by assuming that the height of the population follows a normal distribution. This implies that few people have a shorter or taller height than average.
- **Collecting the sample:** We then collect a sample of heights from the population and find the average height based on that sample.
- **Calculating the likelihood function:** Given the population's average height, we look at the likelihood of observing heights. The likelihood function represents the probability of observing the provided data given the parameters in our model. In our case, the model's parameters are the normal distribution's mean and standard deviation. Due to computational reasons, the log-likelihood function is often used instead of the likelihood function.
- **Maximizing the likelihood function:** MLE aims to find the average height that maximizes the log-likelihood function of obtaining the observed sample and makes the observed heights most probable.



Estimation process
We can now model the average height with a normal distribution whose parameters are selected by maximizing the likelihood function.

Importance of MLE in machine learning

In supervised machine learning, we use labeled data that trains the model's parameters. The training data consists of input features and the corresponding output labels. During the training phase, we aim to find the model parameters that best capture the patterns in the labeled data.

MLE helps fine-tune the machine learning models. In the training phase, we adjust the model's parameters to maximize the likelihood of the labeled data. Alternatively, we can use a negative log-likelihood that represents the loss function. A **loss function** quantifies the difference between predicted and actual values and is defined as follows:

L = |y\{y}^|

Here, y represents the actual output and y^ represents the estimated value. We aim to minimize this loss function L during training to reach an accurate and effective model. Note that minimizing the negative log-likelihood is equivalent to maximizing the likelihood, and this is a common objective in the training of probabilistic models.

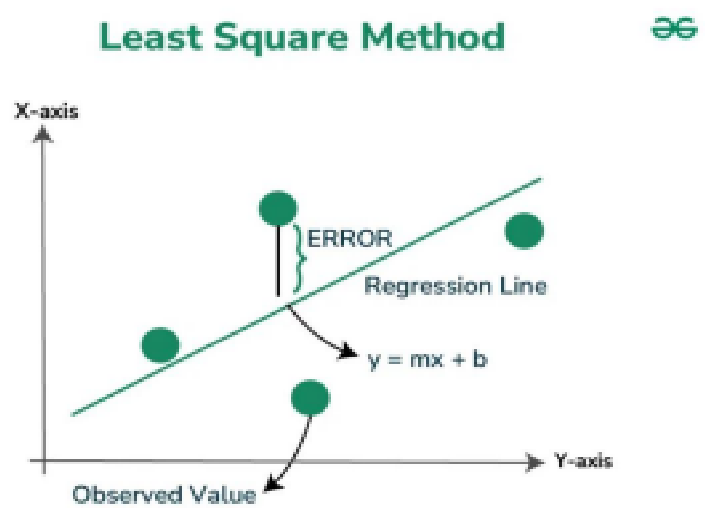
2.Least Square method

Least Square method is a fundamental mathematical technique widely used in **data analysis, statistics, and regression modeling** to identify the **best-fitting curve or line** for a given set of data points. This method ensures that the overall error is reduced, providing a highly accurate model for predicting future data trends.

In statistics, when the data can be represented on a cartesian plane by using the independent and dependent variable as the x and y coordinates, it is called **scatter data.** This data might not be useful in making interpretations or predicting the values of the dependent variable for the independent variable. So, we try to get an **equation of a line that fits best to the given data points** with the help of the **Least Square Method**.

Least Square Method is used to derive a generalized linear equation between two variables. when the value of the dependent and independent variable is represented as the x and y coordinates in a 2D cartesian coordinate system. Initially, known values are marked on a plot. The plot obtained at this point is called a scatter plot. Then, we try to represent all the marked points as a straight line or a **linear equation**. The equation of such a line is obtained with the help of the Least Square method. This is done to get the value of the dependent variable for an independent variable for which the value was initially unknown. This helps us to make predictions for the value of dependent variable.

Least Square Method Definition
Least Squares method is a statistical technique used to find the equation of best-fitting curve or line to a set of data points by minimizing the sum of the squared differences between the observed values and the values predicted by the model.



This method aims at minimizing the sum of squares of deviations as much as possible. The line obtained from such a method is called a **regression line or line of best fit**.

Formula for Least Square Method
Least Square Method formula is used to find the best-fitting line through a set of data points. For a simple linear regression, which is a line of the form y=mx+c, where y is the dependent variable, x is the independent variable, a is the slope of the line, and b is the y-intercept, the formulas to calculate the slope (m) and intercept (c) of the line are derived from the following equations:
1. **Slope (m) Formula:** $m = n(\sum xy)-(\sum x)(\sum y)$ / $n(\sum x2)-(\sum x)2$
2. **Intercept (c) Formula:** $c = (\sum y)-a(\sum x)$ / n
Where:
• n is the number of data points,
• $\sum xy$ is the sum of the product of each pair of x and y values,

• $\sum x$ is the sum of all x values,
• $\sum y$ is the sum of all y values,
• $\sum x^2$ is the sum of the squares of x values.

The steps to find the line of best fit by using the least square method is discussed below: • **Step 1:** Denote the independent variable values as x_i and the dependent ones as y_i. • **Step 2:** Calculate the average values of x_i and y_i as X and Y.

• **Step 3:** Presume the equation of the line of best fit as $y = mx + c$, where m is the slope of the line and c represents the intercept of the line on the Y-axis.
• **Step 4:** The slope m can be calculated from the following formula:

$$m = [\Sigma (X - x_i) \times (Y - y_i)] / \Sigma (X - x_i)^2$$

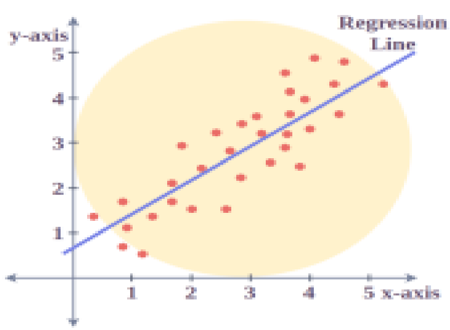
• **Step 5:** The intercept c is calculated from the following formula:

$$c = Y - mX$$

Thus, we obtain the line of best fit as $y = mx + c$, where values of m and c can be calculated from the formulae defined above.
These formulas are used to calculate the parameters of the line that best fits the data according to the criterion of the least squares, minimizing the sum of the squared differences between the observed values and the values predicted by the linear model.

Least Square Method Graph

Let us have a look at how the data points and the line of best fit obtained from the Least Square method look when plotted on a graph.



The red points in the above plot represent the data points for the sample data available. **Independent variables are plotted as x-coordinates and dependent ones are plotted as y-coordinates**. The equation of the line of best fit obtained from the Least Square method is plotted as the red line in the graph.

We can conclude from the above graph that how the **Least Square method helps us to find a line that best fits the given data points** and hence can be used to make further predictions about the value of the dependent variable where it is not known initially. **Limitations of the Least Square Method**

The Least Square method assumes that the data is evenly distributed and doesn't contain any outliers for deriving a line of best fit. But, this method doesn't provide

accurate results for unevenly distributed data or for data containing outliers.
Check: Least Square Regression Line

Least Square Method Solved Examples
Problem 1: Find the line of best fit for the following data points using the Least Square method: (x,y) = (1,3), (2,4), (4,8), (6,10), (8,15).
Solution:
Here, we have x as the independent variable and y as the dependent variable. First, we calculate the means of x and y values denoted by X and Y respectively.
X = (1+2+4+6+8)/5 = 4.2
Y = (3+4+8+10+15)/5 = 8

3.Robust linear regression

Robust linear regression is a type of regression analysis designed to overcome the limitations of ordinary least squares (OLS) regression when the data contains outliers or is not normally distributed. Unlike OLS, which minimizes the sum of squared residuals, robust linear regression minimizes a different function of the residuals to reduce the influence of outliers on the regression model.

Why Use Robust Linear Regression?

1. **Outliers**: OLS regression is highly sensitive to outliers. A few extreme values can significantly affect the estimated coefficients.
2. **Non-normal errors**: If the residuals are not normally distributed, OLS estimations might be biased or inefficient.
3. **Data with heavy tails**: Data that do not follow a normal distribution (heavy-tailed distributions) can lead to unreliable OLS estimates.

Applications

• Robust regression is useful in real-world applications where data may not be perfectly clean, such as in economics, engineering, finance, and environmental science. • It is particularly useful when you expect a high level of noise or outliers in the data that could distort the results from standard linear regression methods.

Summary

Robust linear regression methods are essential when dealing with real-world data that may contain outliers or is not normally distributed. By using robust methods like Huber Regression, RANSAC, or LAD, you can obtain more reliable estimates in the presence of outliers or non normal errors.