

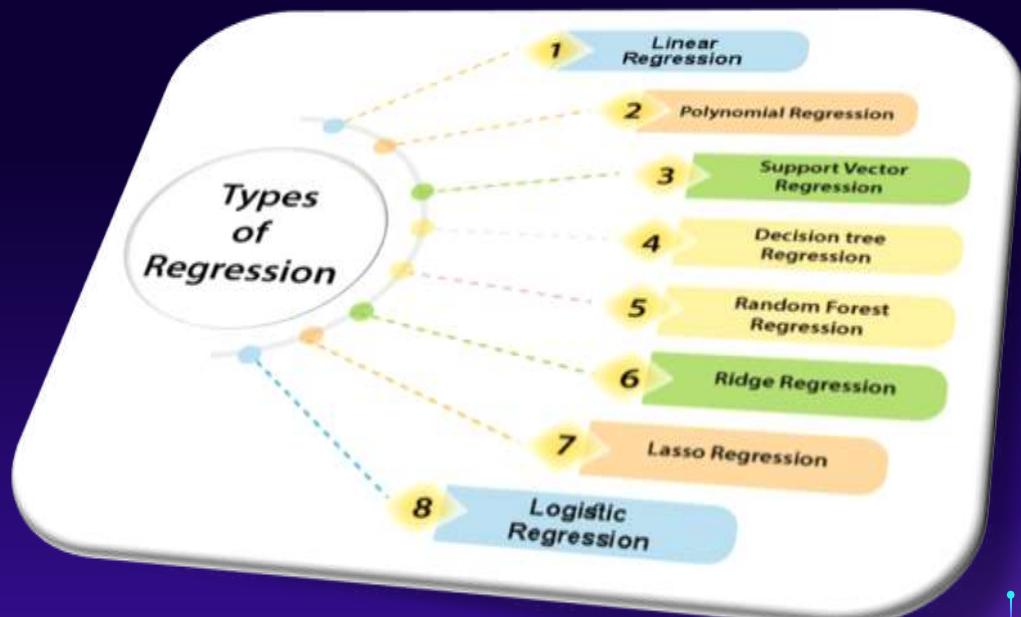
Linear Regression

What is Regression?

Regression is a supervised learning technique which helps in finding the correlation between variables and enables us to predict the continuous output variable based on the one or more predictor variables.

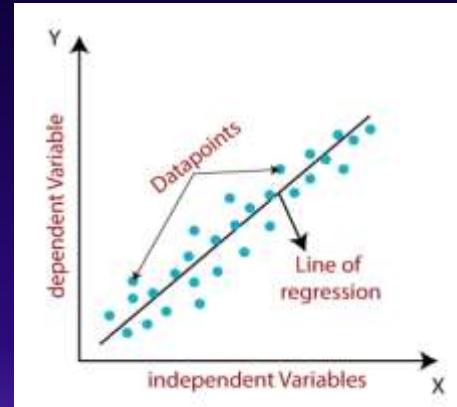
It is mainly used for prediction, forecasting, time series modeling, and determining the causal-effect relationship between variables. Regression analysis helps in the prediction of a continuous variable.

Types of Regression



What is Linear Regression?

- Linear regression analysis is used to predict the value of a variable based on the value of another variable
- Linear regression is a type of supervised machine learning algorithm that computes the linear relationship between a dependent variable and one or more independent features.

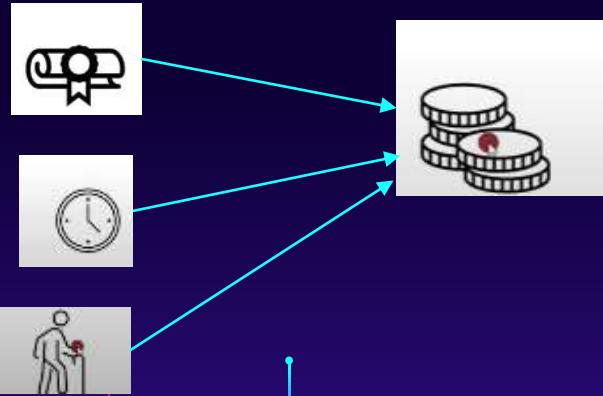


Types of linear regression

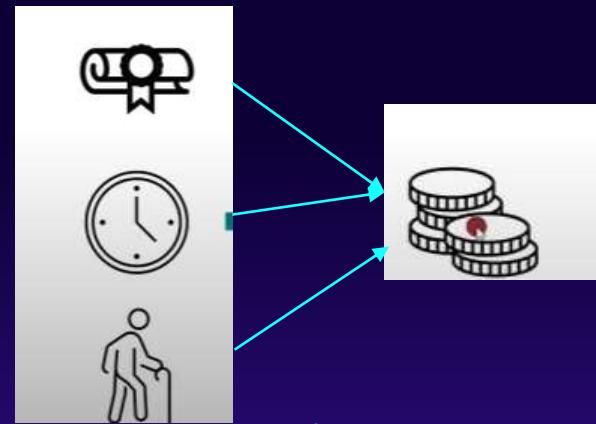
- **Simple linear regression :** The goal of a simple linear regression is to predict the value of a dependent variable based on an independent variable. Simple linear regression is used to model the relationship between two continuous variables. Often, the objective is to predict the value of an output variable (or response) based on the value of an input (or predictor) variable.
- **Multiple linear regression:** Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable.

Types of linear regression

Simple linear regression



Multiple linear regression



Equation of Simple linear regression

$$Y = mX + b$$

- Y represents the dependent variable
- X represents the independent variable
- m is the slope of the line(how much Y changes for a unit change in X)
- b is the intercept (the value of Y when X is 0)

Example : Predicting Pizza Prices



Diameter (X) in Inches	Price(Y) in Dollars
8	10
10	13
12	16

What will be the 20-inch pizza price?



Example : Predicting Pizza Prices



Diameter(X) In Inches	Price(Y) In Dollars	Mean(X)	Mean(Y)	Deviations(X)	Deviations(Y)	Product of Deviations	Sum of Product of Deviations	Square of Deviations for X
8	10	10	13	-2	-3	6	12	4
10	13			0	0	0		0
12	16			2	3	6		4

$m = \text{sum of product of deviations} / \text{sum of square of deviation for } x$

$$= 12/8$$

$$= 1.5$$

$b = \text{Mean of } Y - (m * \text{Mean of } X)$

$$= 13 - (1.5 * 10)$$

$$= -2$$

$$Y = mX + b$$

$$= 1.5X - 2$$

$$= 1.5 * 20 - 2$$

$$= 28 \text{ Dollar}$$



Multiple linear regression formula

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \varepsilon$$

Diagram illustrating the components of the Multiple Linear Regression formula:

- Dependent Variable (Response Variable) points to the term Y .
- Independent Variables (Predictors) point to the terms $\beta_1 X_1, \beta_2 X_2, \dots$.
- Y intercept points to the term β_0 .
- Slope Coefficient points to the term β_1 .
- Error Term points to the term ε .

Example of Multiple Linear Regression

x1 Product 1 Sales	x2 Product 2 Sales	Y Weekly Sales
1	4	1
2	5	6
3	8	8
4	2	12

Here, the Matrices for Y and X are given as follows

$$X = \begin{pmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{pmatrix} \text{ and } Y = \begin{pmatrix} 1 \\ 6 \\ 8 \\ 12 \end{pmatrix}$$

Example of Multiple Linear Regression

The Coefficient of the Multiple Regression Equation is given as

$$\alpha = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

The regression coefficient for multiple regression is calculated the same way as linear regression:

$$\hat{\alpha} = ((X^T X)^{-1} X^T) Y$$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 10 & 19 \\ 10 & 30 & 46 \\ 19 & 46 & 109 \end{pmatrix}$$

Example of Multiple Linear Regression

$$(X^T X)^{-1} = \begin{pmatrix} 4 & 10 & 19 \\ 10 & 30 & 46 \\ 19 & 46 & 109 \end{pmatrix}^{-1} = \begin{pmatrix} 3.15 & -0.59 & -0.30 \\ -0.59 & 0.20 & 0.016 \\ -0.30 & 0.016 & 0.054 \end{pmatrix}$$

The regression coefficient for multiple regression is calculated the same way as linear regression:

$$\hat{a} = ((X^T X)^{-1} X^T) Y$$

$$X^T X)^{-1} X^T = \begin{pmatrix} 3.15 & -0.59 & -0.30 \\ -0.59 & 0.20 & 0.016 \\ -0.30 & 0.016 & 0.054 \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{pmatrix} = \begin{pmatrix} 0.05 & 0.47 & -1.02 & 0.19 \\ -0.32 & -0.098 & 0.155 & 0.26 \\ -0.065 & 0.005 & 0.185 & -0.125 \end{pmatrix}$$

Example of Multiple Linear Regression

$$\hat{a} = ((X^T X)^{-1} X^T) Y = \begin{pmatrix} 0.05 & 0.47 & -1.02 & 0.19 \\ -0.32 & -0.098 & 0.155 & 0.26 \\ -0.065 & 0.005 & 0.185 & -0.125 \end{pmatrix} \times \begin{pmatrix} 1 \\ 6 \\ 8 \\ 12 \end{pmatrix} = \begin{pmatrix} -1.69 \\ 3.48 \\ -0.05 \end{pmatrix}$$

The regression coefficient for multiple regression is calculated the same way as linear regression:

$$\hat{a} = ((X^T X)^{-1} X^T) Y$$

$$a_0 = -1.69$$

$$a_1 = 3.48$$

$$a_2 = -0.05$$

$$y = a_0 + a_1 x_1 + a_2 x_2$$

Hence, the constructed model is:

$$y = -1.69 + 3.48x_1 - 0.05x_2$$

Cost Function in Linear Regression

For the Linear regression model, the cost function will be the minimum of the Root Mean Squared Error of the model, obtained by subtracting the predicted values from actual values. It is a numeric value that indicates the success or failure of a particular model without needing to understand the inner workings of a model.

Why is the cost function needed?

Basically, our machine learning model predicts the new value. We aim to have the predicted value as close as possible to the actual value. If the model predicts the value close to the actual value, then we will say it's a good model. So here is where cost function plays an important role. This cost function is needed to calculate the difference between actual and predicted values. So here it is nothing, just the difference between the actual values-predicted values.

Types of Cost Function in Linear Regression

Mean error: These errors can be negative or positive. Therefore, they can cancel each other out during the summation, and the average error of the model will be zero.

Mean Squared Error : MSE represents the average squared difference between the predictions and expected results.

$$MSE = \frac{1}{2m} \sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2$$

I = index of sample

\hat{Y} = predicted value

Y = expected value

M = number of samples in the data set

Types of Cost Function in Linear Regression

Mean Absolute Error: MSE is a variation of MAE that squares the difference instead of taking the absolute value of the difference. There is no possibility of negative errors.

I = index of sample
 \hat{Y} = predicted value
Y = expected value
M = number of samples in the data set

$$MAE = \frac{1}{m} \sum_{i=1}^m |\hat{y}^{(i)} - y^{(i)}|$$

Types of Cost Function in Linear Regression

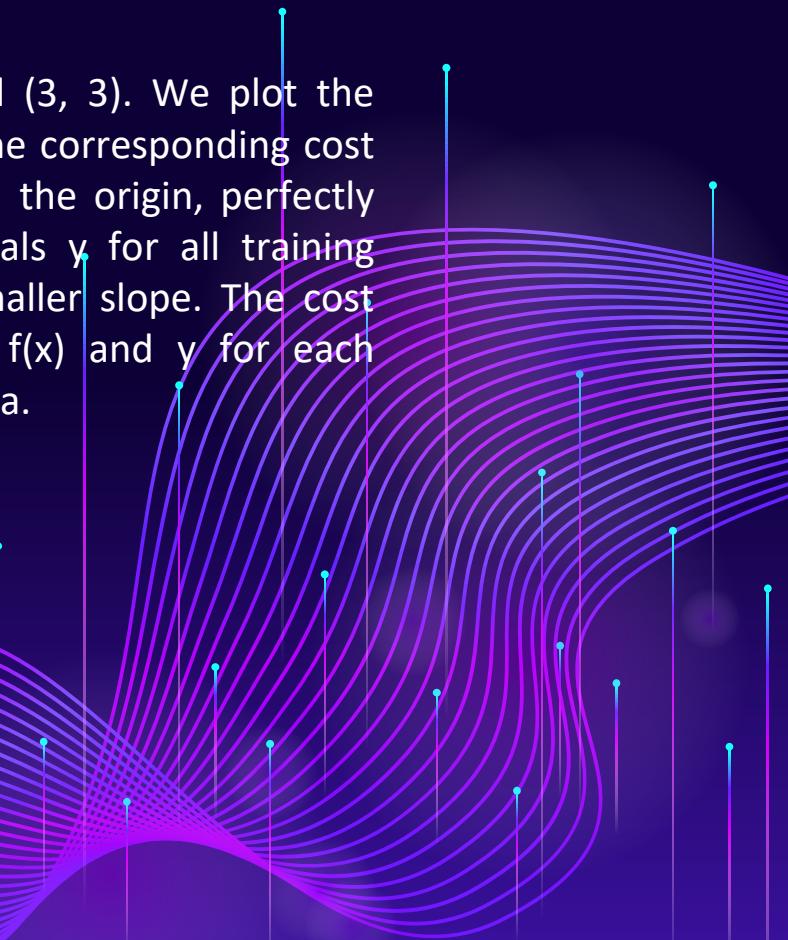
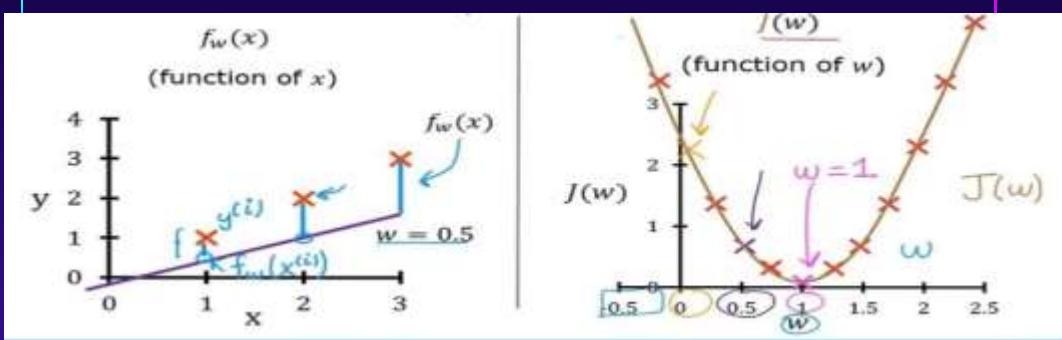
Root mean squared error (RMSE): The Root Mean Squared Error (RMSE) is one of the two main performance indicators for a regression model. It measures the average difference between values predicted by a model and the actual values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

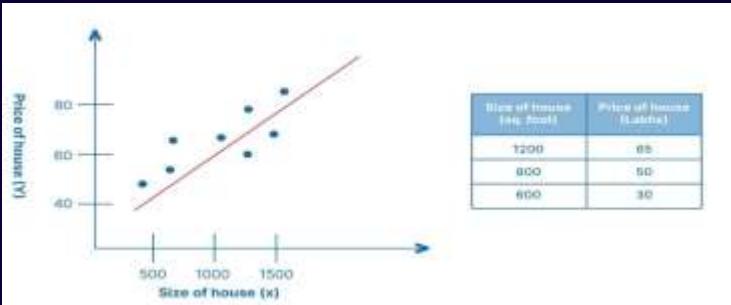
Exploring the Cost Function

We have a training set with three points $(1, 1)$, $(2, 2)$, and $(3, 3)$. We plot the function $f(x) = w * x$ for different values of w and calculate the corresponding cost function $J(w)$. When $w = 1$: $f(x) = x$, the line passes through the origin, perfectly fitting the data. The cost function $J(w)$ is 0 since $f(x)$ equals y for all training examples. Setting $w = 0.5$: $f(x) = 0.5 * x$, the line has a smaller slope. The cost function $J(w)$ now measures the squared errors between $f(x)$ and y for each example. It provides a measure of how well the line fits the data.

Function $f(x)$ vs. Cost Function $J(w)$



Real-life example of cost function



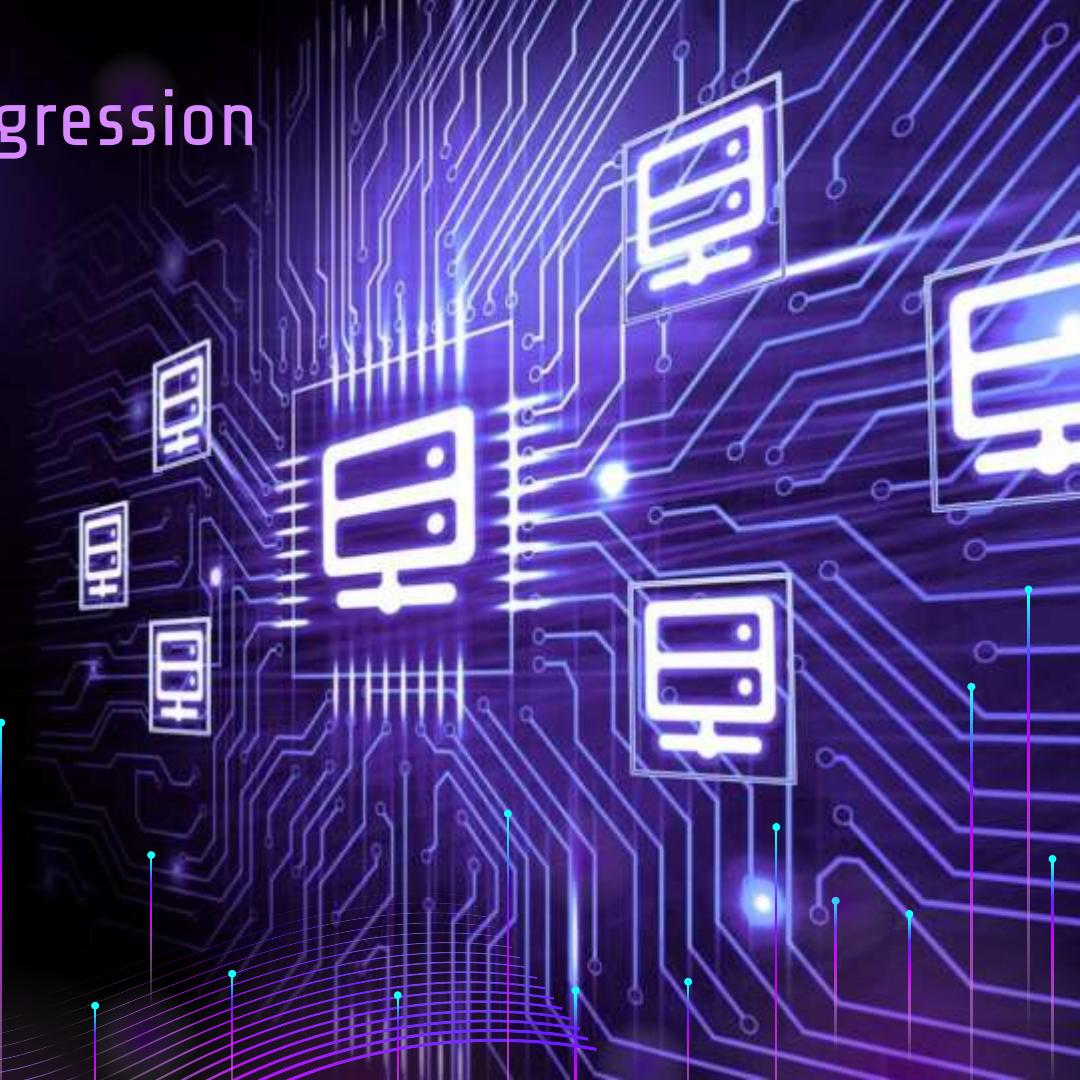
Here, y denotes the target variable(Price of House), X denotes the size of the House and a_0 is the intercept. The linear regression model will fit a straight line to our data, as shown in fig, and it will tell us what is the price of the House when we know the size of the House. Suppose this housing company sells houses having areas under 600 to 1500 square feet. So the prices will vary according to the size of the plot. For example

House(1200 to 1500 square feet)= 65 lakhs.

Now suddenly, there's a rise in demand of Houses which means the price of this House will increase now. Suppose the price has now increased to 75 lakhs. But what if the model predicted its price as 30 lakhs? So this difference between these actual and predicted values is calculated by the cost function.

Applications of Linear Regression

- Market analysis
- Financial analysis
- Sports analysis
- Environmental health
- Medicine



Advantages of Linear Regression

- Linear Regression is simple to implement and easier to interpret the output coefficients
- When you know the relationship between the independent and dependent variable have a linear relationship, this algorithm is the best to use because of it's less complexity compared to other algorithms.
- Linear Regression is susceptible to over-fitting but it can be avoided using some dimensionality reduction techniques, regularization (L1 and L2) techniques and cross-validation.

Drawbacks of Linear Regression

- Linear Assumption
- Sensitivity to Exceptions
- Independence Presumption
- Limited to Linear Relationships

Conclusion

Linear regression is a foundational and widely used technique in machine learning, offering a straightforward yet effective way to model relationships between variables. It is a valuable tool for making predictions and understanding data patterns, though its suitability depends on the specific problem and the linearity assumption. While it may not always be the best choice, it remains an essential component of the machine learning toolkit, providing valuable insights in various domains.

References

- ❖ <https://www.geeksforgeeks.org/ml-advantages-and-disadvantages-of-linear-regression/>
- ❖ <https://iq.opengenus.org/advantages-and-disadvantages-of-linear-regression/>
- ❖ <https://www.intellspot.com/linear-regression-examples/>
- ❖ <https://www.tutorialspoint.com/advantages-and-disadvantages-of-linear-regression#:~:text=The%20strategy%20is%20simple%20to,the%20subordinate%20and%20free%20factors.>
- ❖ https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704-EP713_MultivariableMethods/
- ❖ <https://www.javatpoint.com/simple-linear-regression-in-machine-learning>
- ❖ <https://youtu.be/zUQr6HAAKp4?si=V-ojf6pXnMO2p7DR>
- ❖ <https://youtu.be/lzGKRSvs5HM?si=RduTOfWWjb05n5tF>
- ❖ <https://www.shiksha.com/online-courses/articles/cost-function-in-machine-learning/#Why-is-the-cost-function-needed?>
- ❖ <https://medium.com/@yennhi95zz/3-understanding-the-cost-function-in-linear-regression-for-machine-learning-beginners-ec9edeeecbdde>
- ❖ <https://datatab.net/tutorial/linear-regression>
- ❖ <https://www.youtube.com/watch?v=QcPycBZomac>

Thank You

