

✓ **Machine Learning**

- Machine Learning Tutorial
- Machine Learning Applications
- Life cycle of Machine Learning
- Install Anaconda & Python
- AI vs Machine Learning
- How to Get Datasets
- Data Preprocessing
- Supervised Machine Learning
- Unsupervised Machine Learning
- Supervised vs Unsupervised Learning

✓ **Supervised Learning**

- Regression Analysis
- Linear Regression
- Simple Linear Regression
- Multiple Linear Regression
- Backward Elimination
- Polynomial Regression

✓ **Classification**

- Classification Algorithm
- Logistic Regression
- K-NN Algorithm
- Support Vector Machine Algorithm
- Naïve Bayes Classifier

✓ **Miscellaneous**

- Classification vs Regression
- Linear Regression vs Logistic Regression
- Decision Tree Classification Algorithm
- Random Forest Algorithm
- Clustering in Machine Learning
- Hierarchical Clustering in Machine Learning
- K-Means Clustering Algorithm
- Apriori Algorithm in Machine Learning
- Association Rule Learning
- Confusion Matrix
- Cross-Validation
- Data Science vs Machine Learning
- Machine Learning vs Deep Learning
- Dimensionality Reduction Technique
- Machine Learning Algorithms
- Overfitting & Underfitting
- Principal Component Analysis
- What is P-Value
- Regularization in Machine Learning
- Examples of Machine Learning
- Semi-Supervised Learning
- Essential Mathematics for Machine Learning
- Overfitting in Machine Learning
- Types of Encoding Techniques
- Feature Selection Techniques in Machine Learning
- Bias and Variance in Machine Learning
- Machine Learning Tools
- Prerequisites for Machine Learning
- Gradient Descent in Machine Learning
- Machine Learning Experts Salary India
- Machine Learning Models
- Machine Learning Books
- Linear Algebra for Machine Learning
- Types of Machine Learning
- Feature Engineering for Machine Learning

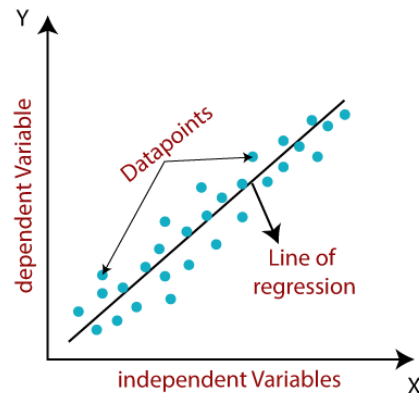
## Linear Regression in Machine Learning

[< Prev](#)[Next >](#)

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as **sales, salary, age, product price**, etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (x) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:



Mathematically, we can represent a linear regression as:

$$y = a_0 + a_1x + \epsilon$$

Here,

Y = Dependent Variable (Target Variable)

X = Independent Variable (predictor Variable)

a<sub>0</sub> = intercept of the line (Gives an additional degree of freedom)

a<sub>1</sub> = Linear regression coefficient (scale factor to each input value).

ε = random error

The values for x and y variables are training datasets for Linear Regression model representation.

## Types of Linear Regression

Linear regression can be further divided into two types of the algorithm:

- **Simple Linear Regression:**

If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

- **Multiple Linear regression:**

If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

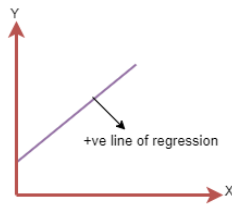
- Top 10 Machine Learning Courses in 2021
- Epoch in Machine Learning
- Machine Learning with Anomaly Detection
- What is Epoch
- Cost Function in Machine Learning
- Bayes Theorem in Machine Learning
- Perceptron in Machine Learning
- Entropy in Machine Learning
- Issues in Machine Learning
- Precision and Recall in Machine Learning
- Genetic Algorithm in Machine Learning
- Normalization in Machine Learning
- Adversarial Machine Learning
- Basic Concepts in Machine Learning
- Machine Learning Techniques
- AutoML
- Demystifying Machine Learning
- Challenges of Machine Learning
- Model Parameter vs Hyperparameter
- Hyperparameters in Machine Learning
- Importance of Machine Learning
- Machine Learning and Cloud Computing
- Anti-Money Laundering using Machine Learning
- Data Science Vs. Machine Learning
- Big Data
- Popular Machine Learning Platforms
- Deep learning vs. Machine learning vs. Artificial Intelligence
- Machine Learning Application in Defense/Military
- Machine Learning Applications in Media
- How can Machine Learning be used with Blockchain
- Prerequisites to Learn Artificial Intelligence and Machine Learning
- List of Machine Learning Companies in India
- Mathematics Courses for Machine Learning
- Probability and Statistics Books for Machine Learning
- Risks of Machine Learning
- Best Laptops for Machine Learning
- Machine Learning in Finance
- Lead Generation using Machine Learning
- Machine Learning and Data Science Certification
- What is Big Data and Machine Learning
- How to Save a Machine Learning Model
- Machine Learning Model with Teachable Machine
- Data Structure for Machine Learning
- Hypothesis in Machine Learning
- Gaussian Discriminant Analysis
- How Machine Learning is used by Famous Companies
- Introduction to Transfer Learning ML
- LDA in Machine Learning
- Stacking in Machine Learning
- CNB Algorithm
- Deploy a Machine Learning Model using Streamlit Library
- Different Types of Methods for Clustering Algorithms in ML

## Linear Regression Line

A linear line showing the relationship between the dependent and independent variables is called a **regression line**. A regression line can show two types of relationship:

- **Positive Linear Relationship:**

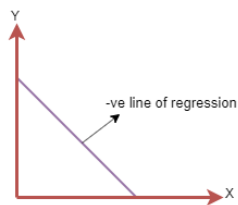
If the dependent variable increases on the Y-axis and independent variable increases on X-axis, then such a relationship is termed as a Positive linear relationship.



The line equation will be:  $Y = a_0 + a_1X$

- **Negative Linear Relationship:**

If the dependent variable decreases on the Y-axis and independent variable increases on the X-axis, then such a relationship is called a negative linear relationship.



The line of equation will be:  $Y = -a_0 + a_1X$

## Finding the best fit line:

When working with linear regression, our main goal is to find the best fit line that means the error between predicted values and actual values should be minimized. The best fit line will have the least error.

The different values for weights or the coefficient of lines ( $a_0$ ,  $a_1$ ) gives a different line of regression, so we need to calculate the best values for  $a_0$  and  $a_1$  to find the best fit line, so to calculate this we use cost function.

### Cost function-

- The different values for weights or coefficient of lines ( $a_0$ ,  $a_1$ ) gives the different line of regression, and the cost function is used to estimate the values of the coefficient for the best fit line.
- Cost function optimizes the regression coefficients or weights. It measures how a linear regression model is performing.
- We can use the cost function to find the accuracy of the **mapping function**, which maps the input variable to the output variable. This mapping function is also known as **Hypothesis function**.

For Linear Regression, we use the **Mean Squared Error (MSE)** cost function, which is the average of squared error occurred between the predicted values and actual values. It can be written as:

For the above linear equation, MSE can be calculated as:

$$MSE = \frac{1}{N} \sum_{i=1}^n (y_i - (a_1x_i + a_0))^2$$

Where,

N = Total number of observation

$y_i$  = Actual value

$(a_1x_i + a_0)$  = Predicted value.

**Residuals:** The distance between the actual value and predicted values is called residual. If the observed points are far from the regression line, then the residual will be high, and so cost function will high. If the scatter points are close to the regression line, then the residual will be small and hence the cost function.

### Gradient Descent:

- Gradient descent is used to minimize the MSE by calculating the gradient of the cost function.
- A regression model uses gradient descent to update the coefficients of the line by reducing the cost function.
- It is done by a random selection of values of coefficient and then iteratively update the values to reach the minimum cost function.

## Model Performance:

The Goodness of fit determines how the line of regression fits the set of observations. The process of finding the best model out of various models is called **optimization**. It can be achieved by below method:

### 1. R-squared method:

- R-squared is a statistical method that determines the goodness of fit.
- It measures the strength of the relationship between the dependent and independent variables on a scale of 0-100%.
- The high value of R-square determines the less difference between the predicted values and actual values and hence represents a good model.
- It is also called a **coefficient of determination**, or **coefficient of multiple determination** for multiple regression.
- It can be calculated from the below formula:

$$R\text{-squared} = \frac{\text{Explained variation}}{\text{Total Variation}}$$

## Assumptions of Linear Regression

Below are some important assumptions of Linear Regression. These are some formal checks while building a Linear Regression model, which ensures to get the best possible result from the given dataset.

- **Linear relationship between the features and target:**  
Linear regression assumes the linear relationship between the dependent and independent variables.
- **Small or no multicollinearity between the features:**  
Multicollinearity means high-correlation between the independent variables. Due to multicollinearity, it may difficult to find the true relationship between the predictors and target variables. Or we can say, it is difficult to determine which predictor variable is affecting the target variable and which is not. So, the model assumes either little or no multicollinearity between the features or independent variables.
- **Homoscedasticity Assumption:**  
Homoscedasticity is a situation when the error term is the same for all the values of independent variables. With homoscedasticity, there should be no clear pattern distribution of data in the scatter plot.
- **Normal distribution of error terms:**  
Linear regression assumes that the error term should follow the normal distribution pattern. If error terms are not normally distributed, then confidence intervals will become either too wide or too narrow, which may cause difficulties in finding coefficients.  
It can be checked using the **q-q plot**. If the plot shows a straight line without any deviation, which means the error is normally distributed.
- **No autocorrelations:**  
The linear regression model assumes no autocorrelation in error terms. If there will be any correlation in the error term, then it will drastically reduce the accuracy of the model. Autocorrelation usually occurs if there is a dependency between residual errors.

- Survival Analysis Using Machine Learning
- Traffic Prediction Using Machine Learning
- t-SNE in Machine Learning
- BERT Language Model
- Federated Learning in Machine Learning
- Deep Parametric Continuous Convolutional Neural Network
- Depth-wise Separable Convolutional Neural Networks
- Need for Data Structures and Algorithms for Deep Learning and Machine Learning
- Geometric Model in Machine Learning
- Machine Learning Prediction
- Scalable Machine Learning
- Credit Score Prediction using Machine Learning
- Extrapolation in Machine Learning
- Image Forgery Detection Using Machine Learning
- Insurance Fraud Detection - Machine Learning
- NPS in Machine Learning
- Sequence Classification- Machine Learning
- EfficientNet: A Breakthrough in Machine Learning Model Architecture
- focl algorithm in Machine Learning
- Gini Index in Machine Learning
- Rainfall Prediction using ML
- Major Kernel Functions in Support Vector Machine
- Bagging Machine Learning
- BERT Applications
- Xtreme: MultiLingual Neural Network
- History of Machine Learning
- Multimodal Transformer Models
- Pruning in Machine Learning
- ResNet: Residual Network
- Gold Price Prediction using Machine Learning
- Dog Breed Classification using Transfer Learning
- Cataract Detection Using Machine Learning
- Placement Prediction Using Machine Learning
- Stock Market prediction using Machine Learning
- How to Check the Accuracy of your Machine Learning Model
- Interpretability and Explainability Transformer Models
- Pattern Recognition in Machine Learning
- Zillow Home Value (Zestimate) Prediction in ML
- Fake News Detection Using Machine Learning
- Genetic Programming VS Machine Learning
- IPL Prediction Using Machine Learning
- Document Classification Using Machine Learning
- Heart Disease Prediction Using Machine Learning
- OCR with Machine Learning
- Air Pollution Prediction Using Machine Learning
- Customer Churn Prediction Using Machine Learning
- Earthquake Prediction Using Machine Learning

✓ Interview Questions

← Prev

Next →

 For Videos Join Our Youtube Channel: [Join Now](#)

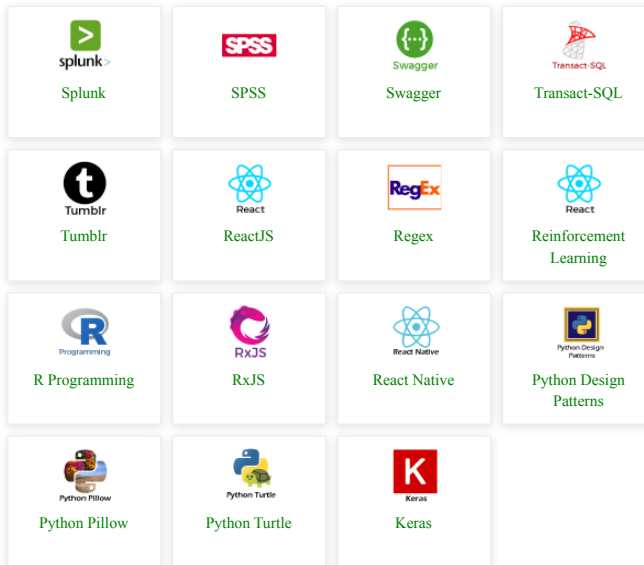
## Feedback

- Send your Feedback to [feedback@javatpoint.com](mailto:feedback@javatpoint.com)

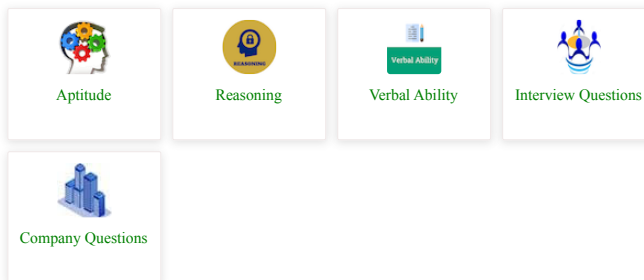
## Help Others, Please Share



## Learn Latest Tutorials

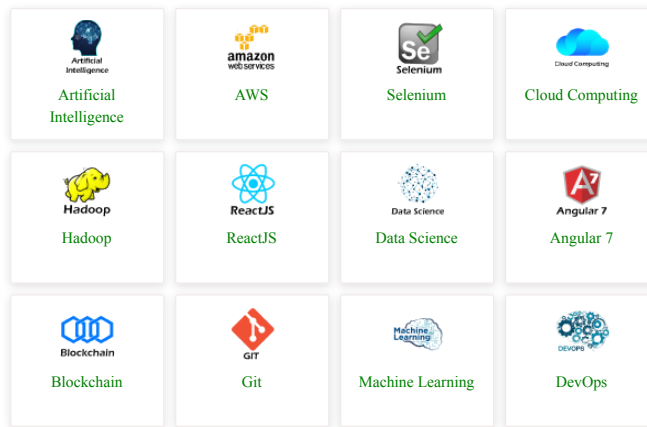


## Preparation

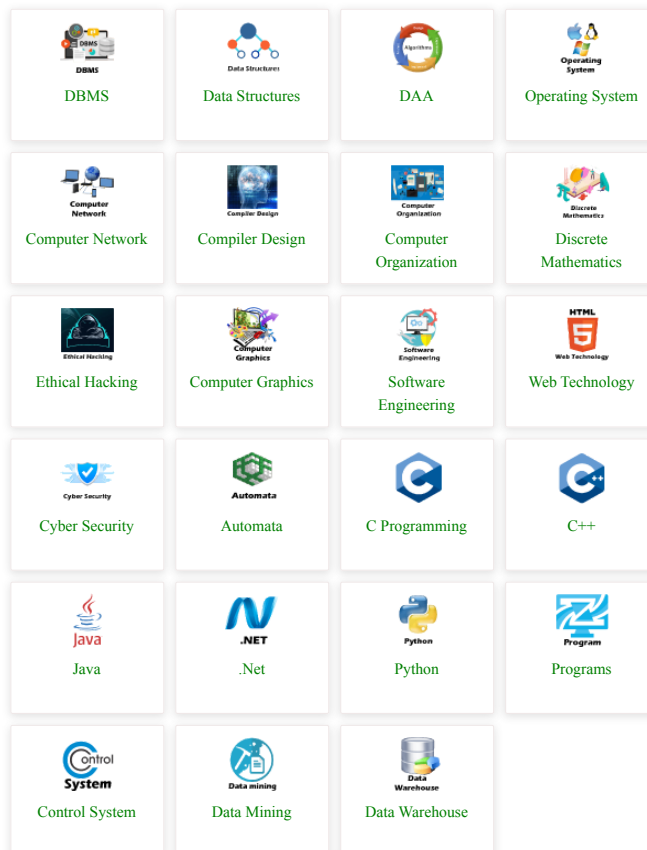


- Factor Analysis in Machine Learning
- Locally Weighted Linear Regression
- Machine Learning in Restaurant Industry
- Machine Learning Methods for Data-Driven Turbulence Modeling
- Predicting Student Dropout Using Machine Learning
- Image Processing Using Machine Learning
- Machine Learning in Banking
- Machine Learning in Education
- Machine Learning in Healthcare
- Machine Learning in Robotics
- Tensorflow Tutorial
- PyTorch Tutorial
- Data Science Tutorial
- AI Tutorial
- NLP Tutorial
- Reinforcement Learning
- Machine learning Interview

## Trending Technologies



## B.Tech / MCA



## Javatpoint Services

JavaTpoint offers too many high quality services. Mail us on [hr@javatpoint.com](mailto:hr@javatpoint.com), to get more information about given services.

- Website Designing
- Website Development
- Java Development
- PHP Development
- WordPress
- Graphic Designing
- Logo
- Digital Marketing
- On Page and Off Page SEO
- PPC
- Content Development
- Corporate Training
- Classroom and Online Training
- Data Entry

## Training For College Campus

JavaTpoint offers college campus training on Core Java, Advance Java, .Net, Android, Hadoop, PHP, Web Technology and Python. Please mail your requirement at [hr@javatpoint.com](mailto:hr@javatpoint.com).

Duration: 1 week to 2 week

Like/Subscribe us for latest updates or newsletter



### LEARN TUTORIALS

Learn Java  
Learn Data Structures  
Learn C Programming  
Learn C++ Tutorial  
Learn C# Tutorial  
Learn PHP Tutorial  
Learn HTML Tutorial  
Learn JavaScript Tutorial  
Learn jQuery Tutorial  
Learn Spring Tutorial

### OUR WEBSITES

[Javatpoint.com](http://Javatpoint.com)  
[Hindi100.com](http://Hindi100.com)  
[Lyricsia.com](http://Lyricsia.com)  
[Quoteperson.com](http://Quoteperson.com)  
[Jobandplacement.com](http://Jobandplacement.com)

### OUR SERVICES

Website Development  
Android Development  
Website Designing  
Digital Marketing  
Summer Training  
Industrial Training  
College Campus Training

### CONTACT

Address: G-13, 2nd Floor, Sec-3  
Noida, UP, 201301, India  
Contact No: 0120-4256464, 9990449935  
Contact Us  
Subscribe Us  
Privacy Policy  
Sitemap  
About Me