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

Regression Analysis in Machine learning

← Prev Next →

Regression analysis is a statistical method to model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables. More specifically, Regression analysis helps us to understand how the value of the dependent variable is changing corresponding to an independent variable when other independent variables are held fixed. It predicts continuous/real values such as **temperature, age, salary, price,** etc.

We can understand the concept of regression analysis using the below example:

Example: Suppose there is a marketing company A, who does various advertisement every year and get sales on that. The below list shows the advertisement made by the company in the last 5 years and the corresponding sales:

Advertisement	Sales
$90	$1000
$120	$1300
$150	$1800
$100	$1200
$130	$1380
$200	??

Now, the company wants to do the advertisement of $200 in the year 2019 **and wants to know the prediction about the sales for this year**. So to solve such type of prediction problems in machine learning, we need regression analysis.

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**.

In Regression, we plot a graph between the variables which best fits the given datapoints, using this plot, the machine learning model can make predictions about the data. In simple words, "***Regression shows a line or curve that passes through all the datapoints on target-predictor graph in such a way that the vertical distance between the datapoints and the regression line is minimum.***" The distance between datapoints and line tells whether a model has captured a strong relationship or not.

Some examples of regression can be as:

- Prediction of rain using temperature and other factors
- Determining Market trends
- Prediction of road accidents due to rash driving.

Terminologies Related to the Regression Analysis:

- **Dependent Variable:** The main factor in Regression analysis which we want to predict or understand is called the dependent variable. It is also called **target variable**.

- Top 10 Machine Learning Courses in 2021
- Epoch in Machine Learning
- Machine Learning with Anomaly Detection
- What is Epoch
- Cost Function in Machine Learnin
- 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 Learnin
- 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 Learni Vs. Big Data
- Popular Machine Learning Platforms
- Deep learning vs. Machine learni vs. Artificial Intelligence
- Machine Learning Application in Defense/Military
- Machine Learning Applications in Media
- How can Machine Learning be us with Blockchain
- Prerequisites to Learn Artificial Intelligence and Machine Learnin
- List of Machine Learning Companies in India
- Mathematics Courses for Machine Learning
- Probability and Statistics Books f Machine Learning
- Risks of Machine Learning
- Best Laptops for Machine Learnir
- 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 Mode using Streamlit Library
- Different Types of Methods for Clustering Algorithms in ML

- **Independent Variable:** The factors which affect the dependent variables or which are used to predict the values of the dependent variables are called independent variable, also called as a **predictor**.
- **Outliers:** Outlier is an observation which contains either very low value or very high value in comparison to other observed values. An outlier may hamper the result, so it should be avoided.
- **Multicollinearity:** If the independent variables are highly correlated with each other than other variables, then such condition is called Multicollinearity. It should not be present in the dataset, because it creates problem while ranking the most affecting variable.
- **Underfitting and Overfitting:** If our algorithm works well with the training dataset but not well with test dataset, then such problem is called **Overfitting**. And if our algorithm does not perform well even with training dataset, then such problem is called **underfitting**.

Why do we use Regression Analysis?

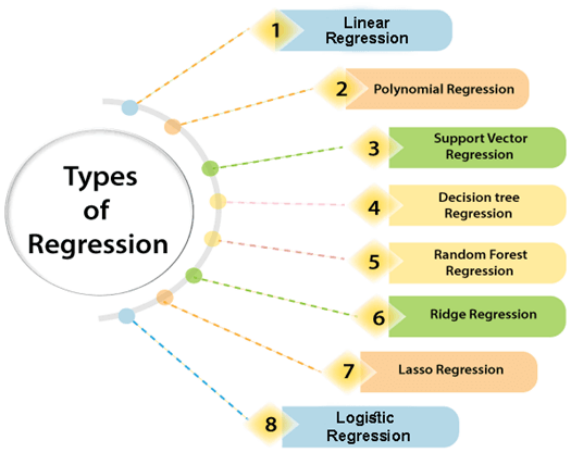
As mentioned above, Regression analysis helps in the prediction of a continuous variable. There are various scenarios in the real world where we need some future predictions such as weather condition, sales prediction, marketing trends, etc., for such case we need some technology which can make predictions more accurately. So for such case we need Regression analysis which is a statistical method and used in machine learning and data science. Below are some other reasons for using Regression analysis:

- Regression estimates the relationship between the target and the independent variable.
- It is used to find the trends in data.
- It helps to predict real/continuous values.
- By performing the regression, we can confidently determine the **most important factor, the least important factor, and how each factor is affecting the other factors**.

Types of Regression

There are various types of regressions which are used in data science and machine learning. Each type has its own importance on different scenarios, but at the core, all the regression methods analyze the effect of the independent variable on dependent variables. Here we are discussing some important types of regression which are given below:

- **Linear Regression**
- **Logistic Regression**
- **Polynomial Regression**
- **Support Vector Regression**
- **Decision Tree Regression**
- **Random Forest Regression**
- **Ridge Regression**
- **Lasso Regression:**

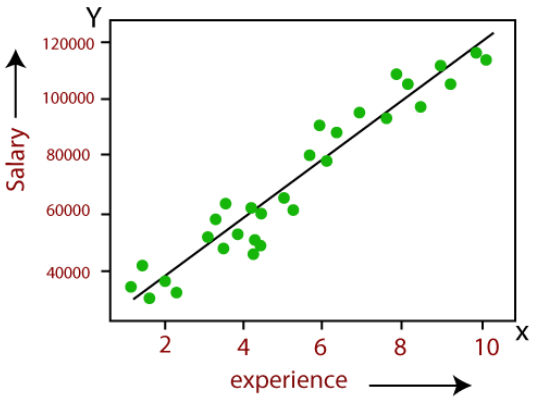


- EM Algorithm in Machine Learnin
- Machine Learning Pipeline
- Exploitation and Exploration in Machine Learning
- Machine Learning for Trading
- Data Augmentation: A Tactic to Improve the Performance of ML
- Difference Between Coding in Da Science and Machine Learning
- Data Labelling in Machine Learni
- Impact of Deep Learning on Personalization
- Major Business Applications of Convolutional Neural Network
- Mini Batch K-means clustering algorithm
- What is Multilevel Modelling
- GBM in Machine Learning
- Back Propagation through time - RNN
- Data Preparation in Machine Learning
- Predictive Maintenance Using Machine Learning
- NLP Analysis of Restaurant Reviews
- What are LSTM Networks
- Performance Metrics in Machine Learning
- Optimization using Hopfield Network
- Data Leakage in Machine Learnir
- Generative Adversarial Network
- Machine Learning for Data Management
- Tensor Processing Units
- Train and Test datasets in Machir Learning
- How to Start with Machine Learning
- AUC-ROC Curve in Machine Learning
- Targeted Advertising using Machi Learning
- Top 10 Machine Learning Project for Beginners using Python
- What is Human-in-the-Loop Machine Learning
- What is MLOps
- K-Medoids clustering-Theoretical Explanation
- Machine Learning Or Software Development: Which is Better
- How does Machine Learning Wor
- How to learn Machine Learning from Scratch
- Is Machine Learning Hard
- Face Recognition in Machine Learning
- Product Recommendation Machir Learning
- Designing a Learning System in Machine Learning
- Recommendation System - Machine Learning
- Customer Segmentation Using Machine Learning
- Detecting Phishing Websites usin Machine Learning
- Hidden Markov Model in Machine Learning
- Sales Prediction Using Machine Learning
- Crop Yield Prediction Using Machine Learning
- Data Visualization in Machine Learning
- ELM in Machine Learning
- Probabilistic Model in Machine Learning

✓ Related Tutorials

Linear Regression:

- Linear regression is a statistical regression method which is used for predictive analysis.
- It is one of the very simple and easy algorithms which works on regression and shows the relationship between the continuous variables.
- It is used for solving the regression problem in machine learning.
- Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), hence called linear regression.
- If there is only one input variable (x), then such linear regression is called **simple linear regression**. And if there is more than one input variable, then such linear regression is called **multiple linear regression**.
- The relationship between variables in the linear regression model can be explained using the below image. Here we are predicting the salary of an employee on the basis of **the year of experience**.



- Below is the mathematical equation for Linear regression:

```
Y= aX+b
```

Here, Y = dependent variables (target variables),
X= Independent variables (predictor variables),
a and b are the linear coefficients

Some popular applications of linear regression are:

- **Analyzing trends and sales estimates**
- **Salary forecasting**
- **Real estate prediction**
- **Arriving at ETAs in traffic.**

Logistic Regression:

- Logistic regression is another supervised learning algorithm which is used to solve the classification problems. In **classification problems**, we have dependent variables in a binary or discrete format such as 0 or 1.
- Logistic regression algorithm works with the categorical variable such as 0 or 1, Yes or No, True or False, Spam or not spam, etc.
- It is a predictive analysis algorithm which works on the concept of probability.
- Logistic regression is a type of regression, but it is different from the linear regression algorithm in the term how they are used.
- Logistic regression uses **sigmoid function** or logistic function which is a complex cost function. This sigmoid function is used to model the data in logistic regression. The function can be represented as:

$$f(x)= \frac{1}{1+e^{-x}}$$

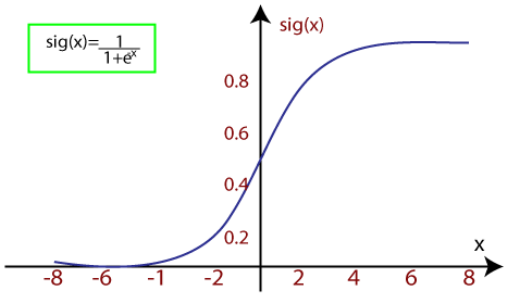
- f(x)= Output between the 0 and 1 value.

- 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 Learnin
- 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 Learnir
- Gini Index in Machine Learning
- Rainfall Prediction using ML
- Major Kernel Functions in Suppor 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 Machi Learning
- Dog Breed Classification using Transfer Learning
- Cataract Detection Using Machin Learning
- Placement Prediction Using Machine Learning
- Stock Market prediction using Machine Learning
- How to Check the Accuracy of yo 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 Machin 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

- x= input to the function
- e= base of natural logarithm.

When we provide the input values (data) to the function, it gives the S-curve as follows:



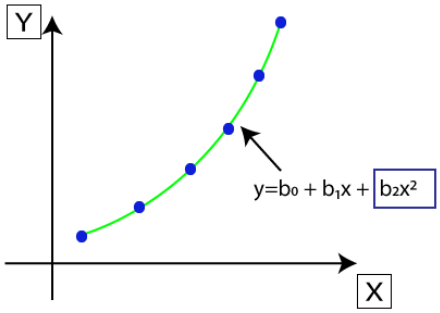
- It uses the concept of threshold levels, values above the threshold level are rounded up to 1, and values below the threshold level are rounded up to 0.

There are three types of logistic regression:

- **Binary(0/1, pass/fail)**
- **Multi(cats, dogs, lions)**
- **Ordinal(low, medium, high)**

Polynomial Regression:

- Polynomial Regression is a type of regression which models the **non-linear dataset** using a linear model.
- It is similar to multiple linear regression, but it fits a non-linear curve between the value of x and corresponding conditional values of y.
- Suppose there is a dataset which consists of datapoints which are present in a non-linear fashion, so for such case, linear regression will not best fit to those datapoints. To cover such datapoints, we need Polynomial regression.
- **In Polynomial regression, the original features are transformed into polynomial features of given degree and then modeled using a linear model.** Which means the datapoints are best fitted using a polynomial line.



- The equation for polynomial regression also derived from linear regression equation that means Linear regression equation $Y= b_0+ b_1x$, is transformed into Polynomial regression equation $Y= b_0+b_1x+ b_2x^2+ b_3x^3+.....+ b_nx^n$.
- Here Y is the **predicted/target output, b_0, b_1,... b_n are the regression coefficients**. x is our **independent/input variable**.
- The model is still linear as the coefficients are still linear with quadratic

> **Note:** This is different from Multiple Linear regression in such a way that in Polynomial regression, a single element has different degrees instead of multiple variables with the same degree.

Support Vector Regression:

Support Vector Machine is a supervised learning algorithm which can be used for regression as well as classification problems. So if we use it for regression problems, then it is termed

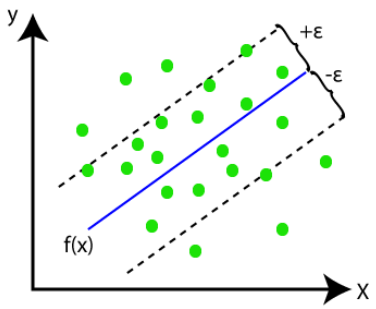
- Factor Analysis in Machine Learning
- Locally Weighted Linear Regressi
- Machine Learning in Restaurant Industry
- Machine Learning Methods for Data-Driven Turbulence Modeling
- Predicting Student Dropout Using Machine Learning
- Image Processing Using Machine Learning
- Machine Learning in Banking
- 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

as Support Vector Regression.

Support Vector Regression is a regression algorithm which works for continuous variables. Below are some keywords which are used in **Support Vector Regression**:

- **Kernel:** It is a function used to map a lower-dimensional data into higher dimensional data.
- **Hyperplane:** In general SVM, it is a separation line between two classes, but in SVR, it is a line which helps to predict the continuous variables and cover most of the datapoints.
- **Boundary line:** Boundary lines are the two lines apart from hyperplane, which creates a margin for datapoints.
- **Support vectors:** Support vectors are the datapoints which are nearest to the hyperplane and opposite class.

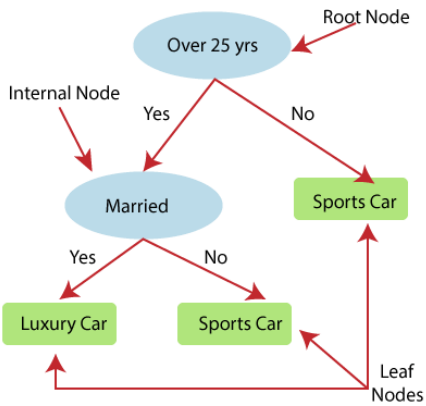
In SVR, we always try to determine a hyperplane with a maximum margin, so that maximum number of datapoints are covered in that margin. ***The main goal of SVR is to consider the maximum datapoints within the boundary lines and the hyperplane (best-fit line) must contain a maximum number of datapoints***. Consider the below image:



Here, the blue line is called hyperplane, and the other two lines are known as boundary lines.

Decision Tree Regression:

- Decision Tree is a supervised learning algorithm which can be used for solving both classification and regression problems.
- It can solve problems for both categorical and numerical data
- Decision Tree regression builds a tree-like structure in which each internal node represents the "test" for an attribute, each branch represent the result of the test, and each leaf node represents the final decision or result.
- A decision tree is constructed starting from the root node/parent node (dataset), which splits into left and right child nodes (subsets of dataset). These child nodes are further divided into their children node, and themselves become the parent node of those nodes. Consider the below image:



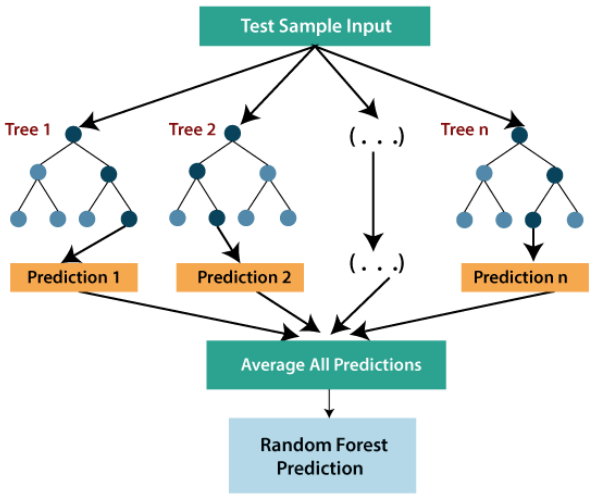
Above image showing the example of Decision Tee regression, here, the model is trying to predict the choice of a person between Sports cars or Luxury car.

- Random forest is one of the most powerful supervised learning algorithms which is capable of performing regression as well as classification tasks.

- The Random Forest regression is an ensemble learning method which combines multiple decision trees and predicts the final output based on the average of each tree output. The combined decision trees are called as base models, and it can be represented more formally as:

```
g(x)= f0(x)+ f1(x)+ f2(x)+....
```

- Random forest uses **Bagging or Bootstrap Aggregation** technique of ensemble learning in which aggregated decision tree runs in parallel and do not interact with each other.
- With the help of Random Forest regression, we can prevent Overfitting in the model by creating random subsets of the dataset.



Ridge Regression:

- Ridge regression is one of the most robust versions of linear regression in which a small amount of bias is introduced so that we can get better long term predictions.
- The amount of bias added to the model is known as **Ridge Regression penalty**. We can compute this penalty term by multiplying with the lambda to the squared weight of each individual features.
- The equation for ridge regression will be:

$$L(x, y) = \text{Min}\left(\sum_{i=1}^{n}(y_i - w_i x_i)^2 + \lambda \sum_{i=1}^{n}(w_i)^2\right)$$

- A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used.
- Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as **L2 regularization**.
- It helps to solve the problems if we have more parameters than samples.

Lasso Regression:

- Lasso regression is another regularization technique to reduce the complexity of the model.
- It is similar to the Ridge Regression except that penalty term contains only the absolute weights instead of a square of weights.
- Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.
- It is also called as **L1 regularization**. The equation for Lasso regression will be:

$$L(x, y) = \text{Min}\left(\sum_{i=1}^{n}(y_i - w_i x_i)^2 + \lambda \sum_{i=1}^{n}|w_i|\right)$$

Next Topic Linear Regression

Feedback

- Send your Feedback to feedback@javatpoint.com

Help Others, Please Share

Learn Latest Tutorials

- Splunk
- SPSS
- Swagger
- Transact-SQL
- Tumblr
- ReactJS
- Regex
- Reinforcement Learning
- R Programming
- RxJS
- React Native
- Python Design Patterns
- Python Pillow
- Python Turtle
- Keras

Preparation

- Aptitude
- Reasoning
- Verbal Ability
- Interview Questions
- Company Questions

Trending Technologies

- Artificial Intelligence
- AWS
- Selenium
- Cloud Computing
- Hadoop
- ReactJS
- Data Science
- Angular 7
- Blockchain
- Git
- Machine Learning
- DevOps

B.Tech / MCA

- DBMS
- Data Structures
- DAA
- Operating System
- Computer Network
- Compiler Design
- Computer Organization
- Discrete Mathematics
- Ethical Hacking
- Computer Graphics
- Software Engineering
- Web Technology
- Cyber Security
- Automata
- C Programming
- C++
- Java
- .Net
- Python
- Programs
- Control System
- Data Mining
- Data Warehouse

Javatpoint Services

JavaTpoint offers too many high quality services. Mail us on 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.
Duration: 1 week to 2 week