

# 10 Popular Regression Algorithms In Machine Learning Of 2022

## INTRODUCTION

Machine Learning (ML) has a wide range of industrial applications that are likely to increase in the coming areas. The global Machine Learning market is expected to reach USD 117 billion by 2027 with an impressive CAGR (Compound Annual Growth Rate) of 39%. Freshers and tech enthusiasts should know about Machine Learning concepts to upskill and build a successful career in the ML industry. Regression algorithms in Machine Learning are an important concept with a lot of use cases.

The future values are predicted with the help of regression algorithms in Machine Learning. The input data/historical data is used to predict a wide range of future values using regression. Label in ML is defined as the target variable (to be predicted) and regression helps in defining the relationship between label and data points. Regression is a type of supervised learning in ML that helps in mapping a predictive relationship between labels and data points. The top types of regression algorithms in ML are linear, polynomial, logistic, stepwise, etc. Read on to know more about the most popular regression algorithms.

## LIST OF REGRESSION ALGORITHMS IN MACHINE LEARNING

1. [Linear Regression](#)
2. [Ridge Regression](#)
3. [Neural Network Regression](#)
4. [Lasso Regression](#)
5. [Decision Tree Regression](#)
6. [Random Forest](#)
7. [KNN Model](#)
8. [Support Vector Machines \(SVM\)](#)
9. [Gaussian Regression](#)
10. [Polynomial Regression](#)

## 1) LINEAR REGRESSION

It is one of the most-used regression algorithms in Machine Learning. A significant variable from the data set is chosen to predict the output variables (future values). Linear regression algorithm is used if the labels are continuous, like the number of flights daily from an airport, etc. The representation of linear regression is  $y = b \cdot x + c$ .

In the above representation, 'y' is the independent variable, whereas 'x' is the dependent variable. When you plot the linear regression, then the slope of the line that provides us the output variables is termed 'b', and 'c' is its intercept. The linear regression algorithms assume that there is a linear relationship between the input and the output. If the dependent and independent variables are not plotted on the same line in linear regression, then there will be a loss in output. The loss in output in linear regression can be calculated as:

Loss function:  $(\text{Predicted output} - \text{actual output})^2$ .

## 2) RIDGE REGRESSION

Ridge Regression is another popularly used linear regression algorithm in Machine Learning. If only one independent variable is being used to predict the output, it will be termed as a linear regression ML algorithm. ML experts prefer Ridge regression as it minimizes the loss encountered in linear regression (discussed above). In place of OLS (Ordinary Least Squares), the output values are predicted by a ridge estimator in ridge regression. The above-discussed linear regression uses OLS to predict the output values.

The complexity of the ML model can also be reduced via ridge regression. One should note that not all the coefficients are reduced in ridge regression, but it reduces the coefficients to a greater extent as compared to other models. The ridge regression is represented as:

$$y = X\beta + \epsilon,$$

where 'y' is the  $N \times 1$  vector defining the observations of the dependent data point/variable and 'X' is the matrix of regressors. ' $\beta$ ' is the  $N \times 1$  vector consisting of regression coefficients and ' $\epsilon$ ' is the vector ( $N \times 1$ ) of errors. The ridge algorithm is also used for regression in Data Mining by IT experts besides ML.

## 3) NEURAL NETWORK REGRESSION

You all must be aware of the power of neural networks in making predictions/assumptions. Each node in a neural network has a respective activation function that defines the output of the node based on a set of inputs. The last activation function can be manipulated to change a neural network into a regression model. One can use 'Keras' that is the appropriate python library for building neural networks in ML.

The output of a neuron is mapped to a variety of values in neural network regression, thus ensuring non-linearity. You can choose a single parameter or a range of parameters for predicting output using neural network regression. The neurons (outputs of a neural network are well-connected with each other, along with a weight associated with each neuron. The well-connected neurons help in predicting future values along with mapping a relationship between dependent and independent variables.

## 4) LASSO REGRESSION

Lasso (Least Absolute Shrinkage and Selection Operator) regression is another widely used linear ML regression (one input variable). The sum of coefficient values is penalized in lasso regression to avoid prediction errors. The determination coefficients in lasso regression are reduced towards zero by using the technique 'shrinkage'. The regression coefficients are reduced by lasso regression to make them fit perfectly with various datasets. Besides ML, the lasso algorithm is also used for regression in Data Mining.

ML experts opt for the lasso regression algorithm when there is high multicollinearity in the given dataset. Multicollinearity in the dataset means independent variables are highly related to each other, and a small change in the data can cause a large change in the regression coefficients. Lasso algorithm regression can be used in predicting forecasting applications in ML.

## 5) DECISION TREE REGRESSION

Non-linear regression in Machine Learning can be done with the help of decision tree regression. The main function of the decision tree regression algorithm is to split the dataset into smaller sets. The subsets of the dataset are created to plot the value of any data point that connects to the problem statement. The splitting of the data set by this algorithm results in a decision tree that has decision and leaf nodes. ML experts prefer this model in cases where there is not enough change in the data set.

One should know that even a slight change in the data can cause a major change in the structure of the subsequent decision tree. One should also not prune the decision tree regressors too much as there will not be enough end nodes left to make the prediction. To have multiple end nodes (regression output values), one should not prune the decision tree regressors excessively.

## 6) RANDOM FOREST

Random forest is also a widely-used algorithm for non-linear regression in Machine Learning. Unlike decision tree regression (single tree), a random forest uses multiple decision trees for predicting the output. Random data points are selected from the given dataset (say  $k$  data points are selected), and a decision tree is built with them via this algorithm. Several decision trees are then modeled that predict the value of any new data point.

Since there are multiple decision trees, multiple output values will be predicted via a random forest algorithm. You have to find the average of all the predicted values for a new data point to compute the final output. The only drawback of using a random forest algorithm is that it requires more input in terms of training. This happens due to the large number of decision trees mapped under this algorithm, as it requires more computational power.

## 7) KNN MODEL

KNN model is popularly used for non-linear regression in Machine Learning. KNN (K Nearest Neighbours) follows an easy implementation approach for non-linear regression in Machine Learning. KNN assumes that the new data point is similar to the existing data points. The new data point is compared to the existing categories and is placed under a relatable category. The average value of the  $k$  nearest neighbors is taken as the input in this algorithm. The neighbors in KNN models are given a particular weight that defines their contribution to the average value.

A common practice of assigning weights to neighbors in a KNN model is  $1/d$ , where  $d$  is the distance of the neighbor from the object whose value is to be predicted. In determining the value of a new data point via the KNN model, one should know that the nearest neighbors will contribute more than the distant neighbors.

## 8) SUPPORT VECTOR MACHINES (SVM)

SVM can be placed under both linear and non-linear types of regression in ML. The use cases of SVM can range from image processing and segmentation, predicting stock market patterns, text categorization, etc. When you have to identify the output in a multidimensional space, the SVM algorithm is used. In a multidimensional space, the data points are not represented as a point in a 2D plot. The data points are represented as a vector in a multidimensional space.

A max-margin hyperplane is created under this model that separates the classes and assigns a value to each class. Freshers should know that an SVM model does not perform to its fullest extent when the dataset has more noise.

## 9. GAUSSIAN REGRESSION

Gaussian **regression algorithms** are commonly used in machine learning applications due to their representation flexibility and inherent uncertainty measures over predictions. A Gaussian process is built on fundamental concepts such as multivariate normal distribution, non-parametric models, kernels, joint and conditional probability.

A Gaussian processes regression (GPR) model can predict using prior knowledge (kernels) and provide uncertainty measures for those predictions. It is a supervised learning method developed by computer science and statistics communities.

Due to the nonparametric nature of Gaussian process regression, it is not constrained by any functional form. As a result, instead of calculating the probability distribution of a specific function's parameters, GPR computes the probability distribution of all permissible functions that fit the data.

## 10. POLYNOMIAL REGRESSION

Polynomial Regression is a regression algorithm that models the relationship between an independent variable ( $x$ ) and a dependent variable ( $y$ ) as an  $n$ th degree polynomial. The equation for Polynomial Regression is as follows:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

It is also known as the special scenario of Multiple Linear Regression in machine learning. Because to make it polynomial regression, some polynomial terms are added to the Multiple Linear Regression equation. It is a linear model that has been modified to improve accuracy. The dataset used for training in polynomial regression is non-linear. To fit the non-linear and complicated functions and datasets. The original features are changed into Polynomial features of the required degree (2,3,...,n) and then modelled using a linear model.

## CONCLUSION

These were some of the top algorithms used for regression analysis. Fresher or not, you should also be aware of all the types of regression analysis. You should also identify the number of variables you are going to use for making predictions in ML. If you have to use only one independent variable for prediction, then opt for a linear regression algorithm in ML.