

## UNIT-5

### REGRESSION AND APPLICATIONS

#### REGRESSION:

Regression refers to a data mining technique that is used to predict the numeric values in a given data set.

Regression can predict all the dependent data sets, expressed in the expression of independent variables, and the trend is available for a finite period. Regression provides a good way to predict variables, but there are certain restrictions and assumptions like the independence of the variables, inherent normal distributions of the variables.

For example, regression might be used to predict the product or service cost or other variables. It is also used in various industries for business and marketing behavior, trend analysis, and financial forecast. In this tutorial, we will understand the concept of regression, types of regression with certain examples.

#### TYPES OF REGRESSION

1. Linear Regression
2. Logistic Regression
3. Lasso Regression
4. Ridge Regression
5. Polynomial Regression

#### LINEAR REGRESSION

Linear regression is the type of regression that forms a relationship between the target variable and one or more independent variables utilizing a straight line. The given equation represents the equation of linear regression

$$Y = a + b * X + e.$$

Where,

a represents the intercept

b represents the slope of the regression line

e represents the error

## **POLYNOMIAL REGRESSION**

If the power of the independent variable is more than 1 in the regression equation, it is termed a polynomial equation. With the help of the example given below, we will understand the concept of polynomial regression.

$$Y = a + b * x^2$$

In the particular regression, the best fit line is not considered a straight line like a linear equation; however, it represents a curve fitted to all the data points.

Applying linear regression techniques can lead to overfitting when you are tempted to minimize your errors by making the curve more complex. Therefore, always try to fit the curve by generalizing it to the issue.

## **LOGISTIC REGRESSION**

When the dependent variable is binary in nature, i.e., 0 and 1, true or false, success or failure, the logistic regression technique comes into existence. Here, the target value (Y) ranges from 0 to 1, and it is primarily used for classification-based problems. Unlike linear regression, it does not need any independent and dependent variables to have a linear relationship.

## **RIDGE REGRESSION**

Ridge regression refers to a process that is used to analyze various regression data that have the issue of multicollinearity. Multicollinearity is the existence of a linear correlation between two independent variables.

Ridge regression exists when the least square estimates are the least biased with high variance, so they are quite different from the real value. However, by adding a degree of bias to the estimated regression value, the errors are reduced by applying ridge regression.

## **LASSO REGRESSION**

The term LASSO stands for Least Absolute Shrinkage and Selection Operator. Lasso regression is a linear type of regression that utilizes shrinkage. In Lasso regression, all the data points are shrunk towards a central point, also known as the mean. The lasso process is most fitted for simple and sparse models with fewer parameters than other regression. This type of regression is well fitted for models that suffer from multicollinearity.

## **APPLICATION OF REGRESSION**

Regression is a very popular technique, and it has wide applications in businesses and industries. The regression procedure involves the predictor variable and response variable. The major application of regression is given below.

- Environmental modeling
- Analyzing Business and marketing behavior
- Financial predictors or forecasting
- Analyzing the new trends and patterns.