# LINEAR REGRESSION

## ***What is a Regression***

*In Regression, we plot a graph between the variables which best fit the given data points. The machine learning model can deliver predictions regarding the data. In other words,****“Regression shows a line or curve that passes through all the data points on a target-predictor graph in such a way that the vertical distance between the data points and the regression line is minimum.”****It is used principally for prediction, forecasting, time series modeling, and determining the causal-effect relationship between variables.*

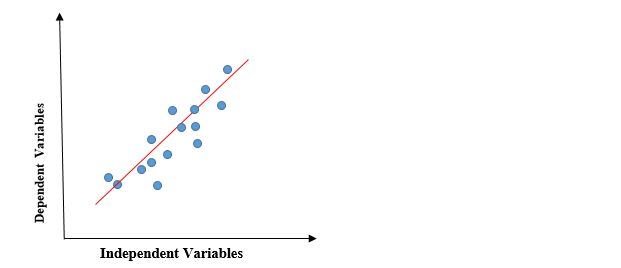
## ***Types of Regression models***

1. *Linear Regression*
2. *Polynomial Regression*
3. *Logistics Regression*

## ***Linear Regression***

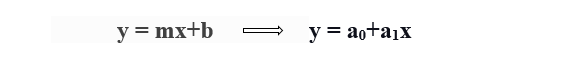
*Linear regression is a quiet and simple statistical regression method used for predictive analysis and shows the relationship between the continuous variables. Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), consequently called linear regression.*

* *If there is a single input variable (x), such linear regression is called******simple linear regression******.*
* *And if there is more than one input variable, such linear regression is called******multiple linear regression****.***
* *The linear regression model gives a sloped straight line describing the relationship within the variable.*

**

* *The above graph presents the linear relationship between the dependent variable and independent variables. When the value of x (****independent variable****) increases, the value of y (****dependent variable****) is likewise increasing. The red line is referred to as the best fit straight line. Based on the given data points, we try to plot a line that models the points the best.*
* ***Calculation of Linear Regression:***

*To calculate best-fit line linear regression uses a traditional slope-intercept form.*

**

* ***y= Dependent Variable.  
    
  x= Independent Variable.  
    
  a0= intercept of the line.  
    
  a1 = Linear regression coefficient.***
* *The goal of the linear regression algorithm is to get the best values for a0 and a1 to find the best fit line. The best fit line should have the least error means the error between predicted values and actual values should be minimized.*
* ***Why Linear Regression:***

*As we discussed, linear regression gives the relationship between dependent variable and independent variable.*

*For example:salary of an employee based on years of experience*

*Here, salary = dependent variable*

*Experience=independent variable*

*By using linear regression we will have the relation between salary and experience i.e, how the salary of an employee is going to vary based on years of experience.*

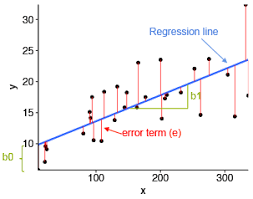
*Now by using the relation we can predict the future salary of the employee.*

* *There are many real time applications like predicting population,weather forecasting,how fast infectious diseases going to spread,etc.*
* ***Types of Linear Regression:***

1. *Simple Linear Regression*
2. *Multiple Linear Regression*

* ***Simple Linear Regression****:*

*In simple linear regression the number of independent variables is one i.e, the regression model explains the relation between single independent and dependent variable.*

**

*Y=mx+c+e*

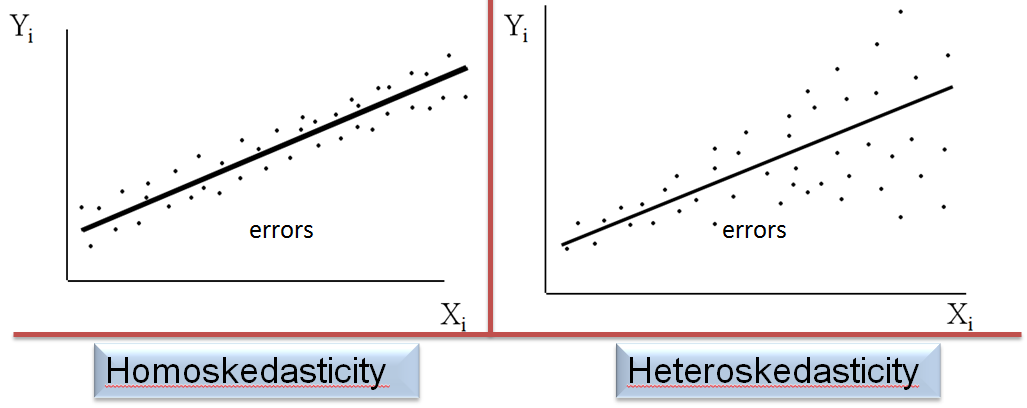
*m=slope of y for a given x*

*c=intercept of y*

*e=error of relation between x and y*

* ***Assumptions of simple Linear Regression:***

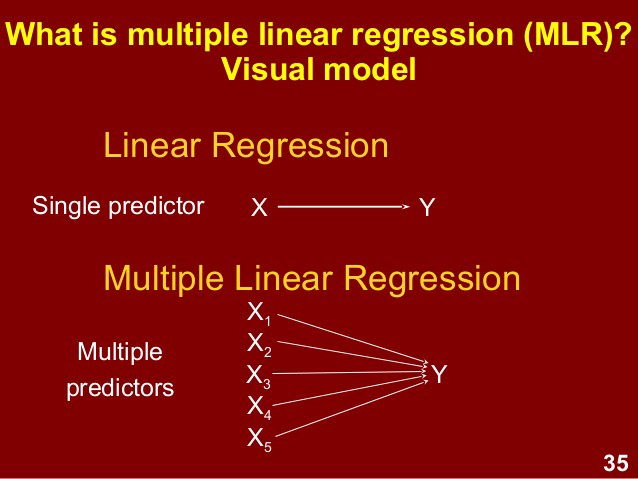
1. *****Homogeneity of variance (homoskedasticity)-******One of the main predictions in a simple linear regression method is that the size of the error stays constant. This simply means that in the value of the independent variable, the error size never changes significantly.*

**

1. *****Independence of observations-******All the relationships between the observations are transparent, which means that nothing is hidden, and only valid sampling methods are used during the collection of data.*
2. *****Normality-******There is a normal rate of flow in the data.*
3. *****The line is always a straight line******-* *There is no curve or grouping factor during the conduction of a linear regression. There is a linear relationship between the variables (dependent variable and independent variable).*

* ***Multiple Linear Regression(MLR):***

Multiple Linear Regression is one of the important regression algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable.



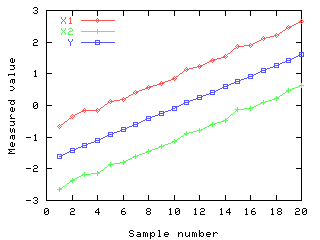
*****Some key points about MLR*:****

* *For MLR, the dependent or target variable(Y) must be the continuous/real, but the predictor or independent variable may be of continuous or categorical form.*
* *Each feature variable must model the linear relationship with the dependent variable.*

The equation for MLR can be represented as:

Y=b0x0+b1x1+b2x2+b3x3+….+bnxn

Where

******

***Y= Output/Response variable***

***b0, b1, b2, b3 , bn....= Coefficients of the model.***

***x1, x2, x3, x4,...= Various Independent/feature variable***

### *Assumptions for Multiple Linear Regression:*

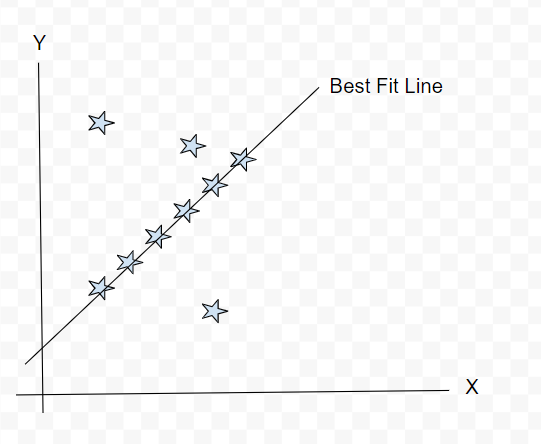
* *A******linear relationship******should exist between the Target and predictor variables.*
* *The regression residuals must be******normally distributed******.*
* *MLR assumes little or******no multicollinearity******(correlation between the independent variable) in data.*
* ***Advantages and Disadvantages of Lineaar Regression:***

|  |  |
| --- | --- |
| *****Advantages***** | *****Disadvantages***** |
| *Linear regression performs exceptionally well for linearly separable data* | *The assumption of linearity between dependent and independent variables* |
| *Easier to implement, interpret and efficient to train* | *It is often quite prone to noise and overfitting* |
| *It handles overftting pretty well using dimensionaliy reduction techniques, regularization, and cross-validation* | *Linear regression is quite sensitive to outliers* |
| One more advantage is the extrapolation beyond a specific data set | *It is prone to multicollinearity* |

* *****Mathematical & Geometrical Intuition of Linear Regression:*****

***Step 1: Plot of the Independent & Dependent Variables. Draw the best fit line (Approx.).***

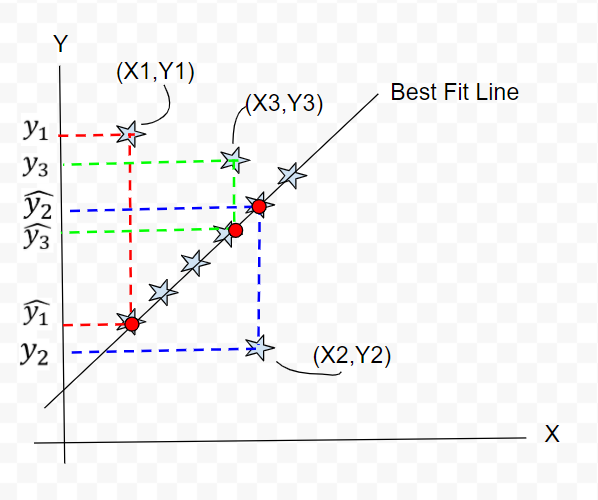
Best Fit Line is determined based on the sum of errors to be minimum.



***Step 2: Calculate the individual errors.***

*Error is defined based on the actual & the predicted value. Below is the formula to calculate the error for individual points.*

*In our Sample Example, we have 3 points which are not in line with our best fit line. So we need to calculate the error function.*

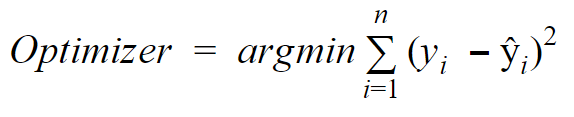


*Error=y-y^*

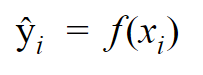
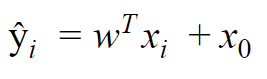
*The calculation of error can conclude the status of selecting the best fit line.*

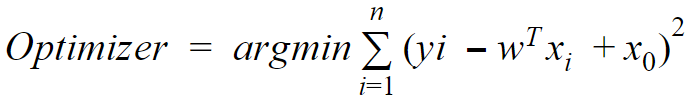
*****Step 3: Calculating the minimum sum of squares of errors or Ordinary Least-squares.*****

*As we have calculated the individual Error for all the points, we are now going to sum & consider the min value to evaluate the best fit line by using the below formula.*



Where,





*Once we have our Best Fit Line on the Linear Model, we are having the freedom to predict the dependent values based on the given independent variables.*

* ***Metrics:***

*There are five error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are:*

1. *Mean Squared Error (MSE).*
2. *Root Mean Squared Error (RMSE).*
3. *Mean Absolute Error (MAE)*
4. *R-squared*
5. *Adjusted R-squared*

## ***Mean Squared Error (MSE)***

*The most common metric for regression tasks is MSE. It has a convex shape. It is the average of the squared difference between the predicted and actual value. Since it is differentiable and has a convex shape, it is easier to optimize.*

IMG_256

*Here,*

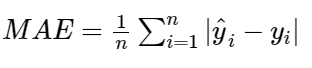
*n=total number of observations*

*yi=Actual value*

*yi^=Predicted value*

## ***Mean Absolute Error (MAE):***

This is simply the average of the absolute difference between the target value and the value predicted by the model. Not preferred in cases where outliers are prominent.



*Here,*

*n=total number of observations*

*yi=Actual value*

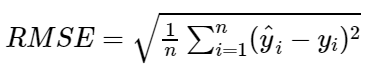
*yi^=Predicted value*

## ***Root Mean Squared Error (RMSE):***

*This is the square root of the average of the squared difference of the predicted and actual value.*

*R-squared error is better than RMSE. This is because R-squared is a relative measure while RMSE is an absolute measure of fit (highly dependent on the variables — not a normalized measure).*

*Basically, RMSE is just the root of the average of squared residuals. We know that residuals are a measure of how distant the points are from the regression line. Thus, RMSE measures the scatter of these residuals.*



*Here,*

*n=total number of observations*

*yi=Actual value*

*yi^=Predicted value*

# ***R-squared:***

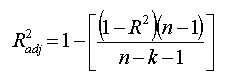
*R-squared finds how much of the variance of the target variable is explained by the model(a function of the independent variables). But to find this we need to know two things.******1) variance of the target variable around mean(avg variance), 2) variance of the target variable around the best-fit line(model variance).*****

IMG_256

1. ***Adjusted R-squared:***

*R2 shows how well terms (data points) fit a curve or line. Adjusted R2 also indicates how well terms fit a curve or line, but adjusts for the number of terms in a model. If you add more and more******useless******[variables](https://www.statisticshowto.com/probability-and-statistics/types-of-variables/)to a model, adjusted r-squared will decrease. If you add more******useful******variables, adjusted r-squared will increase.  
Adjusted R2 will always be less than or equal to R2.*

*You only need R2 when working with****[samples](https://www.statisticshowto.com/sample/)****. In other words, R2 isn’t necessary when you have data from an entire [population](https://www.statisticshowto.com/what-is-a-population/).*

*The formula is:*[](https://www.statisticshowto.com/wp-content/uploads/2013/09/r-squared-adjusted.jpg)  
*where:*

* *N is the number of points in your data sample.*
* *K is the number of independent regressors, i.e. the number of [variables](https://www.statisticshowto.com/probability-and-statistics/types-of-variables/) in your model, excluding the [constant](https://www.calculushowto.com/constant-term-definition/" \t "https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/adjusted-r2/_blank).*
* ***Gradient Descent:***

*An algorithm to minimize the function and by optimizing its parameters. A gradient descent start with a random guess and slowly predict the correct answer. In gradient descent the predictions will go downwards and upwards and this how the parameters are optimized.*

***Types of Gradient descent:***

***1.Batch Gradient descent:***

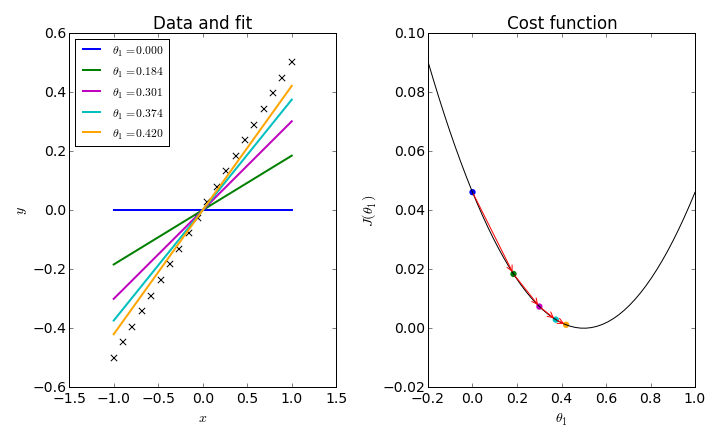
*Batch gradient descent, also called vanilla gradient descent, calculates the error for each example within the training dataset, but only after all training examples have been evaluated does the model get updated. This whole process is like a cycle and it's called a training epoch.*

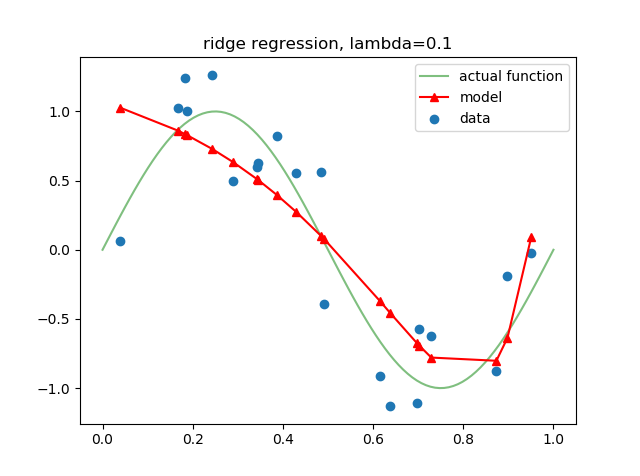
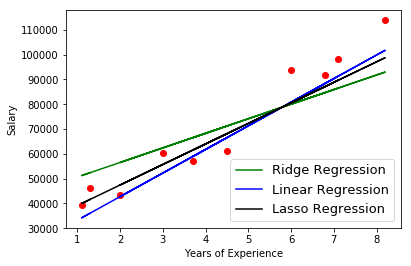
***2.Stochastic Gradient descent:***

*By contrast, stochastic gradient descent (SGD) does this for each training example within the dataset, meaning it updates the parameters for each training example one by one. Depending on the problem, this can make SGD faster than batch gradient descent. One advantage is the frequent updates allow us to have a pretty detailed rate of improvement.*

***3.Mini-Batch Gradient descent:***

*Mini-batch gradient descent is the go-to method since it’s a combination of the concepts of SGD and batch gradient descent. It simply splits the training dataset into small batches and performs an update for each of those batches. This creates a balance between the robustness of stochastic gradient descent and the efficiency of batch gradient descent.*

**

* ***Ridge regression:***
* *Ridge regression is****a method of estimating the coefficients of multiple-regression models in scenarios where linearly independent variables are highly correlated****.*
* *It is most suitable when a data set contains a higher number of predictor(independent) variables than the number of observations. The second-best scenario is when multicollinearity is experienced in a set.*
* *Multicollinearity happens when predictor variables exhibit a correlation among themselves.*
* *Ridge regression aims at reducing the standard error by adding some bias in the estimates of the [regression](https://corporatefinanceinstitute.com/resources/knowledge/finance/regression-analysis/" \o ").*
* *The reduction of the standard error in regression estimates significantly increases the reliability of the estimate.*
* **
* ***Lasso regression:***
* *Lasso regression is****a type of linear regression that uses shrinkage****. Shrinkage is where data values are shrunk towards a central point, like the mean*.
* *This type of regression is used when the dataset shows high multicollinearity or when you want to automate variable elimination and feature selection.*
* *when any features slope value is nearer to zero it will remove that feature considering it as unimportant feature.*
* **