**What is Linear Regression?:**

Linear regression is used for finding linear relationship between target and one or more predictors. There are two types of linear regression- Simple and Multiple.

**Simple Linear Regression:**

Simple linear regression is useful for finding relationship between two continuous variables. One is predictor or independent variable and other is response or dependent variable.

It looks for statistical relationship but not deterministic relationship. Relationship between two variables is said to be deterministic if one variable can be accurately expressed by the other. For example, using temperature in degree Celsius it is possible to accurately predict Fahrenheit. Statistical relationship is not accurate in determining relationship between two variables. For example, relationship between height and weight.

**Question: What is best fit line**

The core idea is to obtain a line that best fits the data. The best fit line is the one for which total prediction error (all data points) are as small as possible.

Error is the distance between the point to the regression line.

Y(predicted op)=m(coificient of slope)\*x(testing input)+c(intercept)

**Question: What is the error function of linear regression**

Error is the distance between the point to the regression line.

The values m and c must be chosen so that they minimize the error. If sum of squared error is taken as a metric to evaluate the model, then goal to obtain a line that best reduces the error.

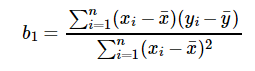


For model with one predictor,

**Intercept calculation**

**b0=c ,b1=m**

**Co-efficient Formula**



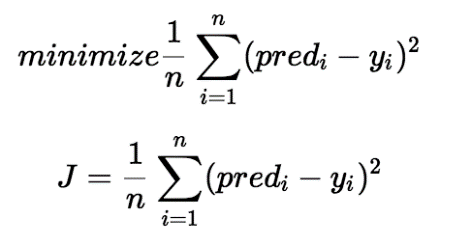
m:

* If m > 0, then x(predictor) and y(target) have a positive relationship. That is increase in x will increase y.
* If m < 0, then x(predictor) and y(target) have a negative relationship. That is increase in x will decrease y.

**Exploring ‘c’**

* If the model does not include x=0, then the prediction will become meaningless with only b0. For example, we have a dataset that relates height(x) and weight(y). Taking x=0(that is height as 0), will make equation have only b0 value which is completely meaningless as in real-time height and weight can never be zero. This resulted due to considering the model values beyond its scope.

The cost function helps us to figure out the best possible values for ‘c’ and ‘m’ which would provide the best fit line for the data points. Since we want the best values for ‘c’ and ‘m’, we convert this search problem into a minimization problem where we would like to minimize the error between the predicted value and the actual value.



[Introduction to Machine Learning Algorithms: Linear Regression | by Rohith Gandhi | Towards Data Science](https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a)

# Coefficient of Determination-R2 score

of determination also called as R2 score is used to evaluate the performance of a linear regression model. It is the amount of the variation in the output dependent attribute which is predictable from the input independent variable(s). It is used to check how well-observed results are reproduced by the model, depending on the ratio of total deviation of results described by the model.

**R2= 1- SSres / SStot**

A higher value of R2 is desirable as it indicates better results.

* The best possible score is 1 which is obtained when the predicted values are the same as the actual values.
* R2 score of baseline model is 0.
* During the worse cases, **R2** score can even be negative.
* Multiple Linear Regression: It’s a form of linear regression that is used when there are two or more predictors.