1. What is Linear Regression?

When ever we start read Algoritham of machine learning, then we start with Linear regression algoritham.

Linear Regression

Best fit Line

↓ use [By doing it]

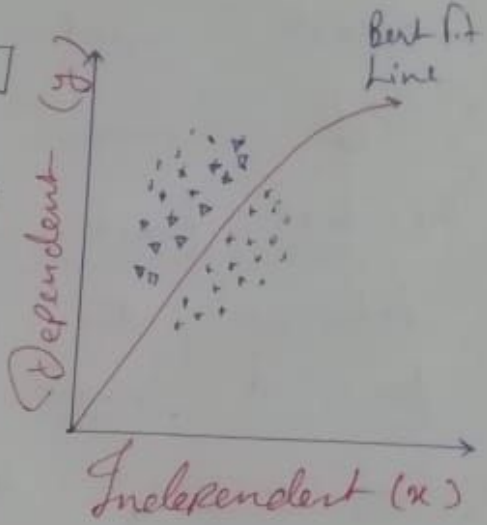
independent and dependent] Variable

↓ in Between

Establish relation

↓

It's called linear Regression.



Equation → $y = mx + c$ - With the help of this equation we find the Best fit Line.

(2) How we Can Calculate Error in Linear Regression

Linear regression most often uses mean - square error (MSE) to calculate the error of the model.

$$MSE \Rightarrow \frac{1}{m} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

n = The total number of terms for which the error is to be calculated.

y_i = The observed value of the variable

$\hat{y}_i$ = The predicted value of the variable

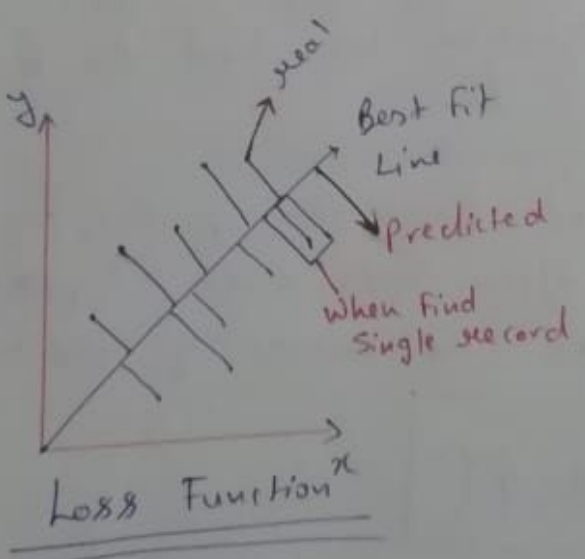
(3) Error is the different between the actual value and predicted value.

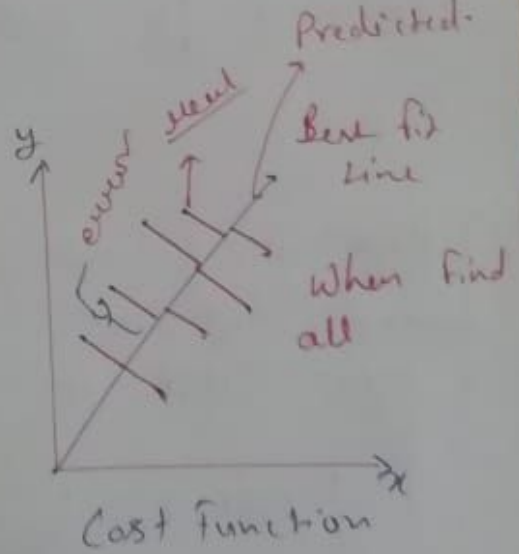
- The real (actual) value is the value derived from observation are measurement of the available data.
- The expected value is the predicted value of the variable based on the regression analysis.

(4) Diff. Between Loss and Cost function?)

The loss function → Is to capture the different between the actual and predicted value for a single record

Cost function → Aggregate the difference for the entire training data set





Formula →

$$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^i - y^{(i)} \right)^2$$

Loss Function

Cost Function

$[h_\theta(x_i) = \theta_0 + \theta_1 x]$

Explane how a ...

MAE Mean absolute Error ⇒ The mean absolute error measure the average different between Predicted values and actual value

MAE ⇒ $\frac{1}{n} \sum_{i=1}^{n} |y - \hat{y}|$

Advantage - ① Reduct to outliers.

② It will also be in the same unit.

DisAdvange → ① Convergen usually take more time optimization in a complex task.

② Time Consuming

MSE and MAE = unlike the mean squared error the MAE calculates the error on the same scale as the data. This mean it's in easier to interpet.

② The MAE don't square the differencies and is less susceptible to outliers.

RMSE = MSE and RMSE are same only different between RMSE is a square root of the MSE

in RMSE also value b/w 0 to 100 %

if RMSE value is more then the model will more good.

in which we determine coefficient.

RMSE → $\sqrt{\frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2}$

⇒ 1 - $\frac{\text{Sum Squared regression error}}{\text{Sum Squared total error}}$

5 MAE, MSE and RMSE ?

A MAE [Mean Square Error] → It's tell us, How much Close data Point to the best Fit Line.

Mean Square error is the average of the Square of the different between the observed and predicted value of a variable

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (y_i - \bar{y}_i)^2$$

m = The total number of terms for which the error in to be calculated

y_i = observed value

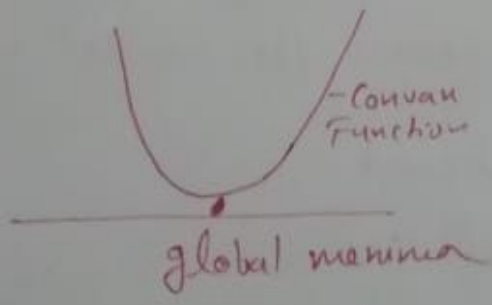
$\bar{y}_i$ = Predicted value.

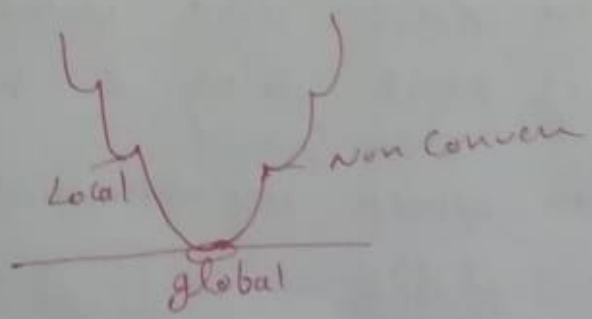
Advantage - ① This equation in differentiable

② This equation also has one global minima.

Disadvantage → ① This is not reduce to outliers.

② Finalizing the error changing the unit





Explain how gradient descent work in Linear regression?

Gradient descent in an iterative optimization algorithm to find the minimum of a function.

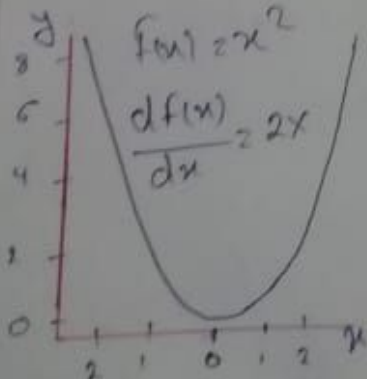
Gradient descent is an iterative o first-order optimization algorithm used to find a local minimum / maximum of a given function.

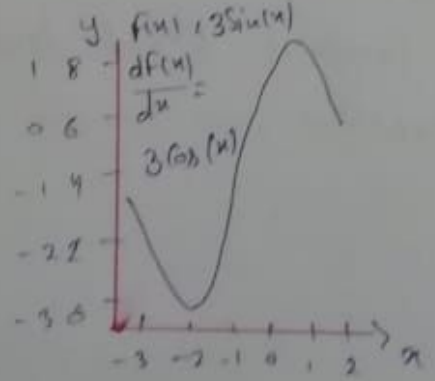
This method in Commonly used in machine learning / and deep learning to minimize a Cost / loss function in a linear Regression.

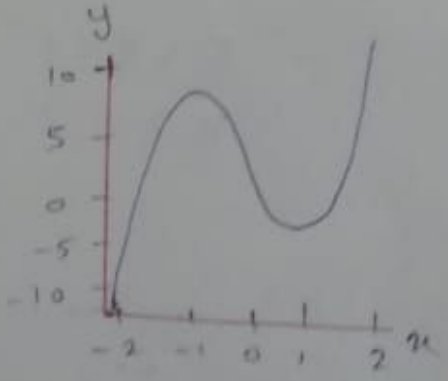
Function requirements → Gradient descent algorithm does not work for all function. There are two specific requirement

- differentiable
- Convex

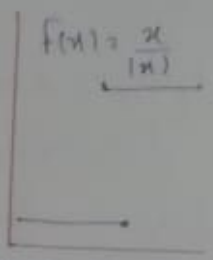
If function is differentiable it has a derivative for each point in its domain. not all function meet these Criteria.





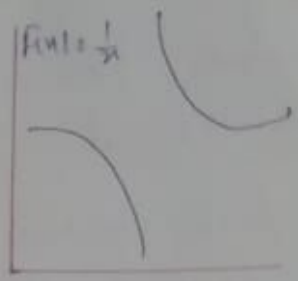


How is hypothesis testing wrong in non differentiable function



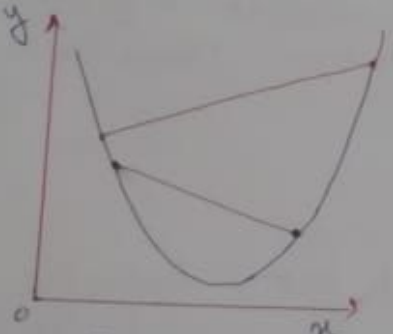
Jump discontinuity

Cusp

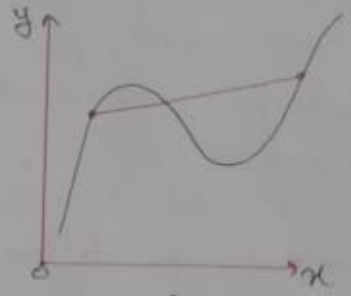


infinite

Convex → This means that the line segment connecting two function points lays on or above it curve (it does not cross it). If it does it mean that it has a local minimum which is not a global one



Convex



Non Convex

for $x < 0$: function is Convex

$0 < x < 1$ Concave (2nd derivative < 0)

$x > 1$: function is Convex again

How is hypothesis testing using in Linear Regression?

Hypothesis testing is used to confirm if our beta coefficients are significant in a linear regression model. Every time we run the linear regression model, we test if the line is significant or not by checking if the coefficient is significant

Hypothesis testing :-

(1) Z - Test → Normal Distribution
- Large Simple Size
- Population Standard deviation

(2) T-test → T Distribution
- Sample Size less than 30
- Population Standard deviation unknown

(3) F - test → Positively Skewed distribution
- when you want to Compare 3 or more Variables

Write all the Assumption for Linear Regression

Assumption for Linear R

Different Between R-Square and Adjusted R-Square ?

R-Square : Residual Sum of Squares.

Adjusted R-Square :- Adjusted Residual Sum of Squares

R-Squared is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variable in a regression model.

Formula :- $R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}}$

$= 1 - \frac{SS_{regression}}{SS_{total}}$

$SS_{regression}$ -> is the sum of squares due to regression

SS_{total} : is the total sum of square.

Adjusted R Square :- The adjusted R-Squared is a modified version of R-Squared that adjusts for predictors that are not significant in a regression model.

$$\text{Adjusted } R^2 = \left\{1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right]\right\}$$

Q) Write all the Assumptions for Linear Regression

- Assumption for Linear Regression.

1) Linear Relationship b/w Independent and dependent Variable.
2) Number of observations should be greater than number of independent variable.
3) No multi - Collinearity in independent Variables.
4) The Variance in the independent Variables should be Positive.
5) Mean of residuals should be zero.
6) No auto - correlation b/w the residuals.
7) Residuals must be normally distributed.
8) Residuals should be constant or equal ~~Variabl~~ Variance .i.e Homoscedasticity.

(7) Explane what the intercept term mean?

The intercept in a regression model represents the mean value of the response variable when all of the predictor variable in the model are equal to zero

Formula

Simple linear Regression => $\hat{y} = \beta_0 + \beta_1(x)$

$\hat{y}$ = Predicted value for the response variable

β_0 = mean value when $x = 0$

β_1 = Average Change in the response variable for a one unit increase in x

x = The value for the predicted value

Multiple Linear Regression =

$$\hat{y} = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_3) + \ldots + \beta K(xK)$$