Decision Tree Algorithm :-

Decision Tree is Supervised Machine Learning Algorithm that can be used for both regression and classification tasks and it is a tree structured classification.

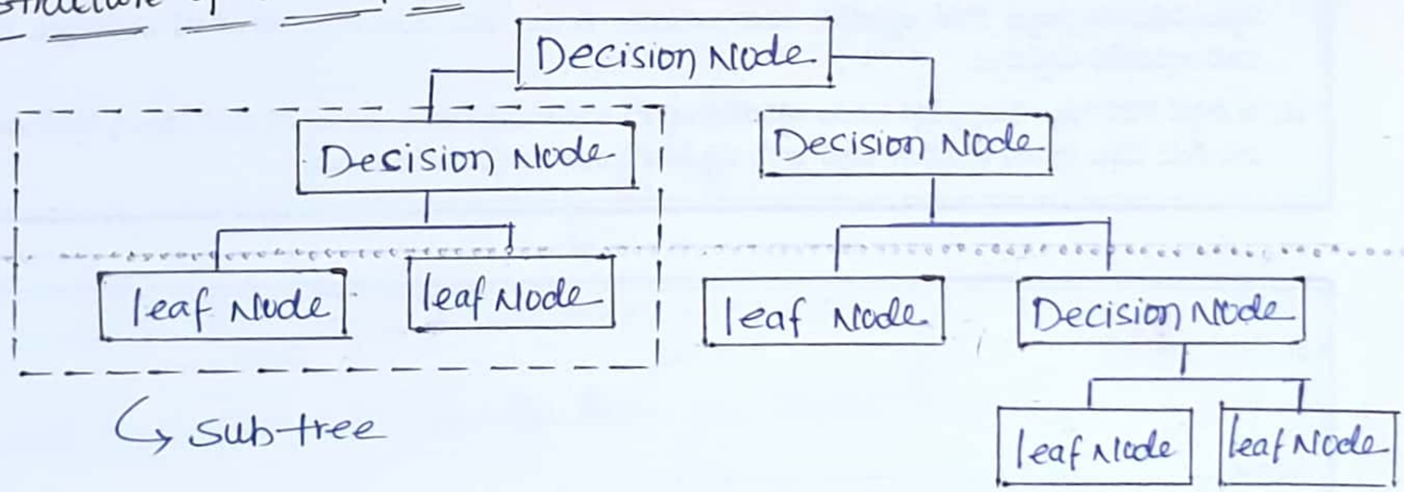
The Internal nodes of a tree represents the features or columns of a dataset, the braches of a tree represents the Decision rules. Each leaf node of a tree represents the Outcome of a descision rule.

In decision tree there are two nodes — ① Decision (or) Root Node
② leaf Node

Decision Node :- Decision node are used to make any decision and have multiple branches.

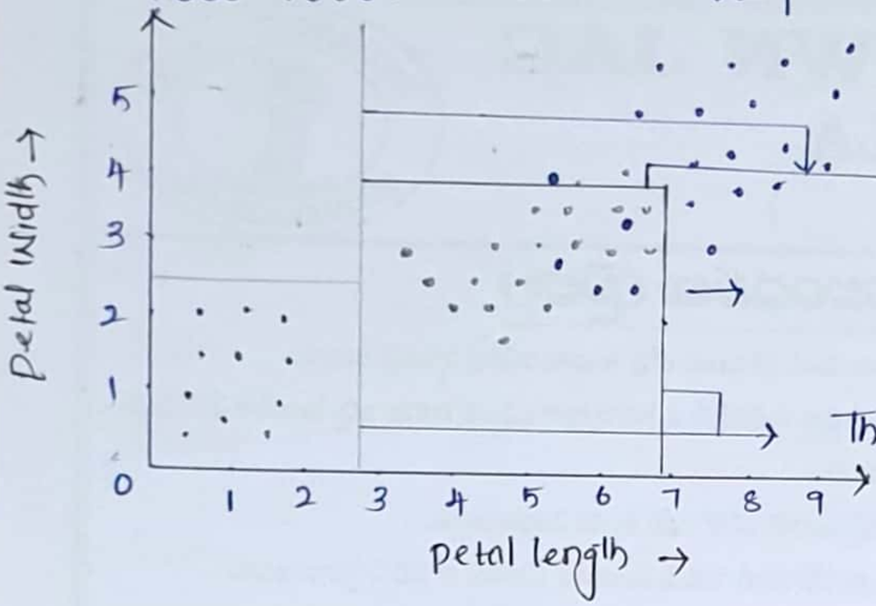
Leaf Node :- leaf Nodes are the Output of a Decision and do not contain any further branches.

Structure of Decision Tree :-



Solving problem by Decision Tree :-

iris_verginica.
iris - versicolor
iris - ~~verginica~~ setosa.



Geometrically -
-Axis parallel Hyperplanes

These lines are Geometrically Represent

if_else Rule (Expert - system) for above Graph :-

According to Programming it is if_else Rule

```
if petal_length < 2.5 :
    point ("setosa")
elif petal_length < 5 and petal_width < 1.7 :
    point ("versicolor")
else :
    point ("verginica")
```

By Geometric Representation :-

-At Geometrical level we are creating multiple lines that can Separate our data, we are just Creating the Boundaries.

In Decision-tree we are categorizing or classifing the data with axis parallel hyperplanes becouse the lines are parallel to axises.

By Creating line we can classify Our data in Geometrical way.

Decision tree used when the data is in Non-linear-form.

Binary Classification Dataset :-

X | Y - Target Vari

Outlook	Temparature	Humidity	Wind	Play Tennis
Sunny	Hot	High	Strong	NO
Overcast	Hot	Normal	Weak	Yes
Rain	Cool	Normal	Strong	NO
Sunny	Cool	Normal	Weak	Yes
Rain	mild	High	Weak	yes
Overcast	Cool	Normal	Strong	yes

→ from this dataset flow chart of Decision Tree :-

- Outlook
  - Sunny → Humidity
    - High → NO
    - Normal → Yes
  - Overcast → Yes
  - Rain → Wind
    - Strong → NO
    - Weak → Yes

Representation of data in form of tree. [Decision Tree]

- Machine learning
  - Supervised ML.
    - Regression
      - Linear
      - K.N.N.
      - D.T
    - classification
      - K.N.N.
      - logistic Reg.
      - Decision Tree → it takes Axis parallel Hyperplane
  - Unsupervised M.L

Gradient Descent :-

Linear Reg — $m^*, c^* = \arg_{m,c} \min \left\{ \sum_{i=1}^{n} (y_i - (mx_i + c))^2 \right\}$

↳ Here we trying to minimize total Squared Error.

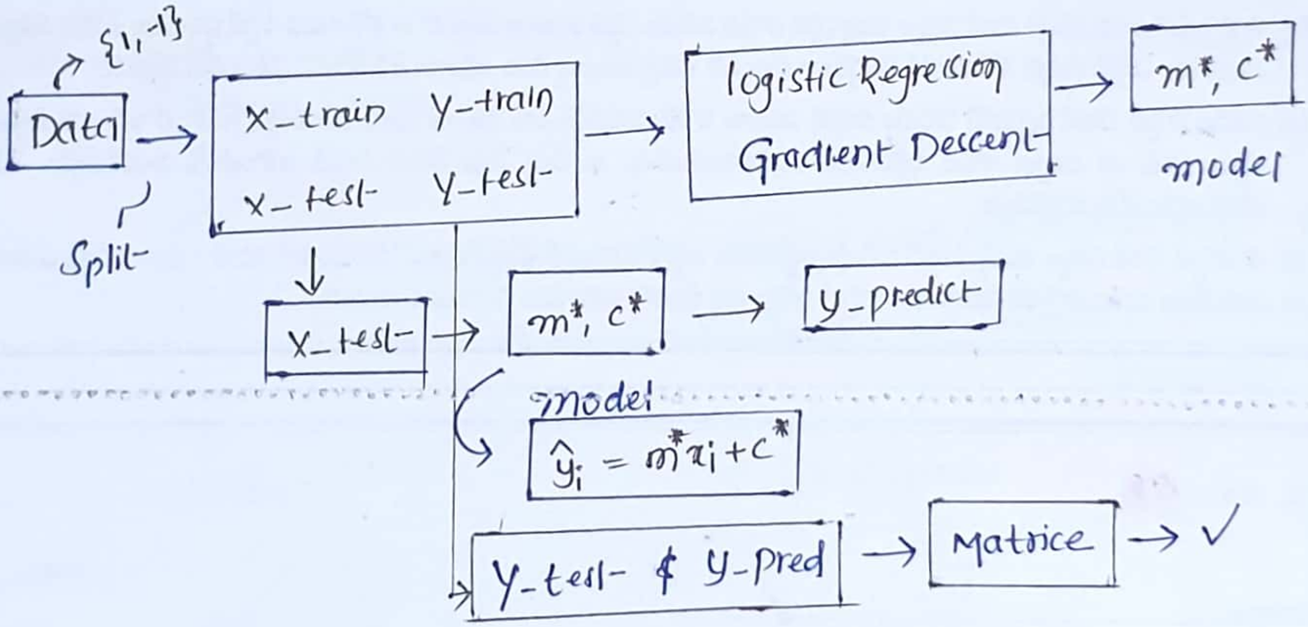
Logistic Reg — $m^*, c^* = \arg_{m,c} \text{Max} \left\{ \sum_{i=1}^{n} \underbrace{(y_i \times (mx_i + c))}_{\text{Signed distance}} \right\}$

↳ Signed distance give us the point Correctly classified or not and how far away that line. In this Sign tells us Correctly classified or not and distance tells us how far away from line

Note : this eqn gets Impacted with the presence of Outlier in the data.

By Using Sigmoid function (σ) we can Solve this problem. (Outlier treatment)

$$\sigma(x) = \frac{1}{1+e^{-x}} \approx \frac{1}{1+\exp\{-x\}}$$



In Above graph we applying Something Called as Gradient descent It will give the Optimal values of m^*, c^*

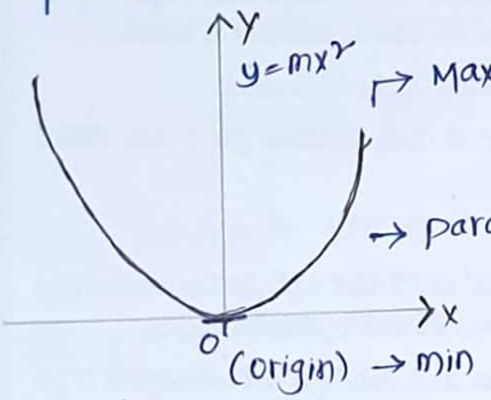
Matrices is Used to evaluting the model performance

without Using gradient descent we can not get the values of m^*, c^* and we Can't Solve the equation.

In All Machine learning Algorithems we will get Optimal values by Using Gradient descent.

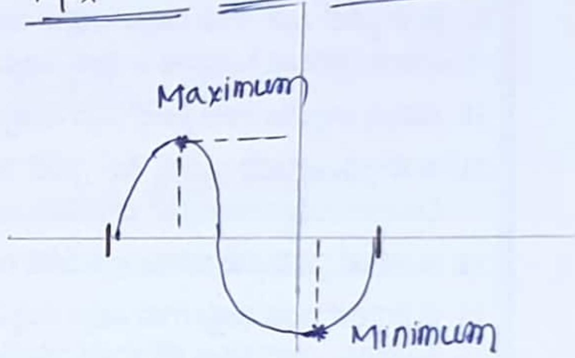
Linear Regression Optimization :- $\Rightarrow \arg_{m,c} \min \left\{ \sum_{i=1}^{n} (y_i - (mx_i + c))^2 \right\}$

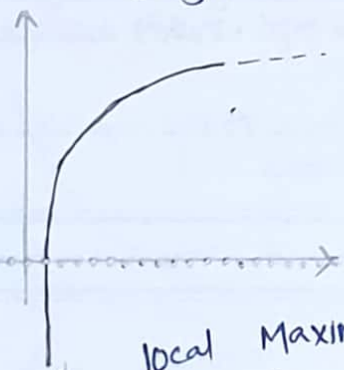
Here we are trying to minimize, and the function is look like Quadratic function ($y = x^2$) becouse of Square and it looks likes in Graph as parabola. and at Origin it will be minimum.



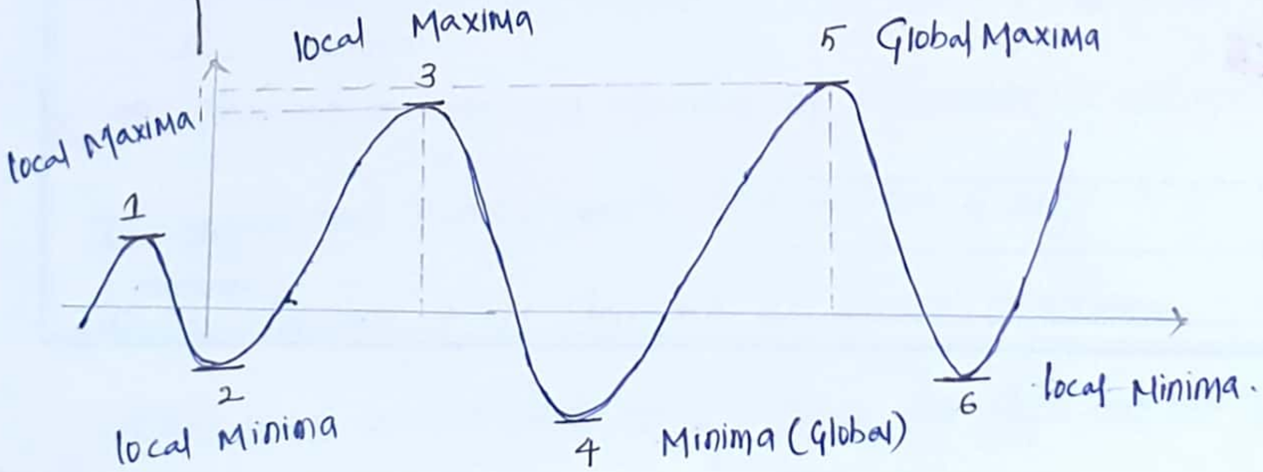
→ parabola function ⇒

Maximum and Minimum function

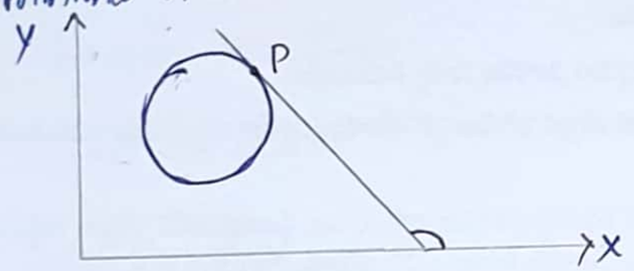




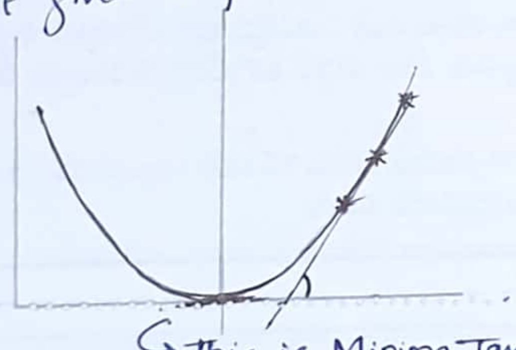
⇒ There is no Maxima and no minima point for This function becouse the function is Continiously increasing.



By the previous function which contain more no of points as maxima, minima values, if we observe Graph there are 5 points In that 2 represent minima and 3 are maxima; if we observe the the maxima highest value which consider as Global maxima because that is highest value on y axis rest of the points consider as local Maxima. , Similarly if we observe minima points the highest one is Global minima rest is local minima as considered.



if we consider circle function In order to find the slope at this function we basically create a tangent and check the angle b/w tangent and x-axis it gives slope at that point p.



→ The slope at this minima is zero because the tangent and x-axis parallel to each other and angle is 0° ie $\tan(0°) = 0$.

↳ This is Minima Tangent

* The slope at minima and Maxima of a function is always zero because The tangent and x-axis parallel to each other & angle is 0°
  whether function → its slope = 0 at minima or Maxima
* Other than minima and Maxima points the slope will not be zero.

Basic Diffrentiation Techniques :-

Eqn – $\boxed{\frac{d}{dx} x^a = a x^{a-1}}$

Ex ① $\frac{d}{dx} x^2 = 2x^{2-1} = 2x$ | $\frac{d}{dx} 5x = 5 \cdot \frac{d}{dx} x \Rightarrow 5 \cdot \frac{d}{dx} x^0 \quad \{x^0 = 1\}$

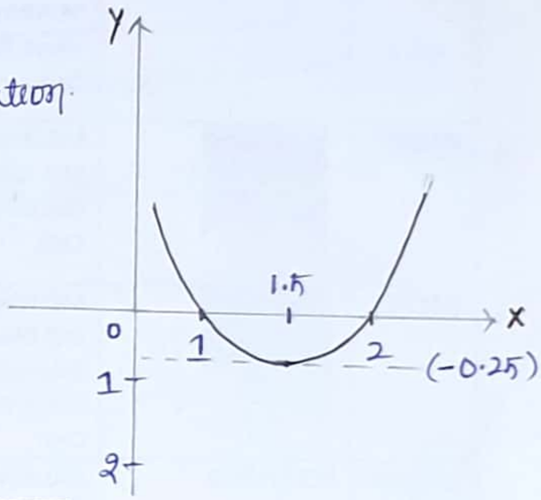
$= 5$

How to Compute the minima of a function $f = x^2 - 3x + 2$

Sol : $f \to \text{Min}\{x^2 - 3x + 2\}$

Computation of minima is nothing but find out the Slope.

We all nothat Slope at minima = '0' zero.



We Can find out Slope by Using diffrentiation.

Now derivation will be – $\frac{df}{dx} = 0$

$$\frac{d}{dx}(x^2 - 3x + 2) = 0$$

$$\Rightarrow \frac{d}{dx} x^2 - \frac{d}{dx} 3x + \frac{d}{dx} 2 = 0$$

$$2x - 3 + 0 = 0 \Rightarrow \boxed{x = \frac{3}{2} \simeq 1.5}$$

∴ at $x = 1.5$ we will get Minima value or Maxima value,

In Order to find wether it will be minima or Maxima we do–

put $x = \frac{3}{2}$ in Eqn $\Rightarrow f \to x^2 - 3x + 2 = 0 \Rightarrow \left(\frac{3}{2}\right)^2 - 3\left(\frac{3}{2}\right) + 2$

$$\frac{9}{4} - \frac{9}{2} + 2 = \boxed{-0.25} \quad \text{Negative value}$$

I dont know the resultant will be minima or maxima.

ie we Cant Say from this value So we have to do Several Observations at $x = 1 \Rightarrow (1)^2 - 3(1) + 2 \Rightarrow 0$

at $x = 2 \Rightarrow (2)^2 - 3(2) + 2 = 4 - 6 + 2 = 0$

Now by Graph representation we can say that at point $x = 1.5$ the function will be minima. and the value is -0.25

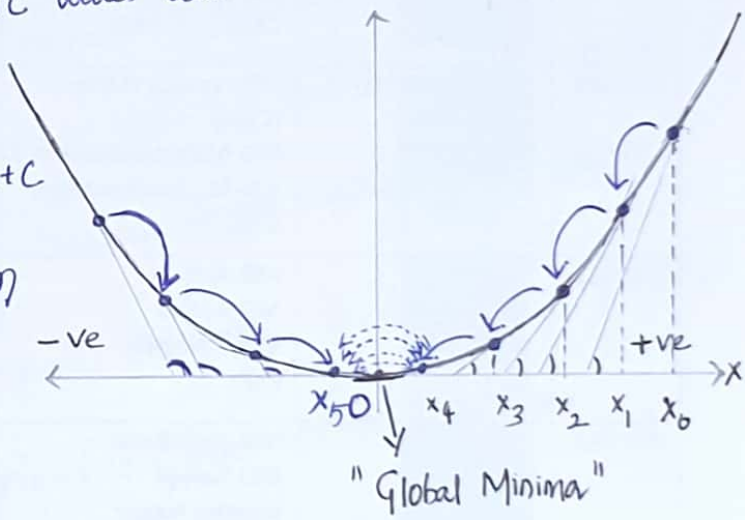
Linear Regression Eqn :- it is nothing but a line that best fit the data which give m^*, c^* value which are optimized values where we have to find out the "Minima".

$$\boxed{m^*, c^* = \arg_{m,c} \text{Min} \left\{ \sum_{i=1}^{n} \left(y_i - (mx_i + c) \right)^2 \right\}}$$

In order to find out the minima we used the Gradient descent algo.

Gradient Descent :- it is an "iterative algorithm" that can be help to find the minima, which help to find m^* \$ c^* values where the function have minima values.

By the above Eqn inform of ax^2+bx+c Quadratic Eqn, Graph representation will be a parabola.



Procedure :-

Randomly Initialize a point either right hand Side or left hand Side, when picking point from right hand positive values comes when ever move to forward because angle is making at that side is $(0-90°)$, when ever from left hand side Negative values comes towards forward moving $(90°-180°)$.

Here we are finding the Slope at initial position x_0 after Calculate Slope we move towards forward direction ie in left hand Side like that moves upto reach the Global minima point

Now we reach to point x_1 towards minima position left Side and generating positive values this Cycle runs till end point (Global minima), when ever move forward we find out Slopes in different locations (like x_0, x_1, x_2 - -)

when ever we move towards forward direction we generate new points. from this we get -

$$x_1 = x_0 - \eta \left[\frac{df}{dx}\right]_{x_0} \text{ (position)}$$

$x_1 \downarrow$ New Generating point; $x_0 \downarrow$ Initial value

where - η - Learning Rate

(η) learning Rate will decide how longer move in forward direction.

Continiously we have to move and find the Slopes (x_1, x_2, x_3 - - - etc)

If we observe graph after point (x_5) turn towards right hand Side direction ie opposite to previous jumps then we calculate Slope at the point x_5 We get Negative value because angle making to that point is lying in between ($90° - 180°$), finally Oscilating in between x_4 to x_5 points repeatedly Untill the end point (ie Global Minima) reaches.

$$\boxed{X_{new} = X_{old} - \eta * \left[\frac{d}{dx} f\right]_{X_{old}}}$$