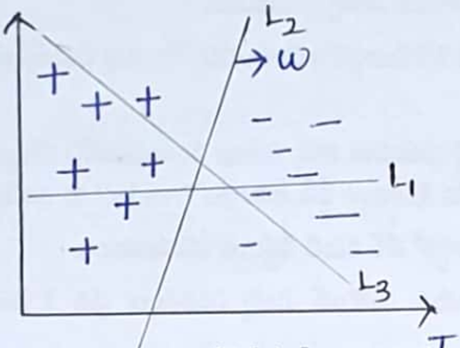
Logistic Regression :-

it is a Supervised learning classification algorithm used to Predict the probability of target variable it is one of Simplest ML-Algorithm that can be Used for various Classification problems Such that Spam detection, Cancer detection etc.

In logistic Regression Our task is to find a line that BEST SEPARATE +ve's from -ve's — $\boxed{D_n = \{(x_i, y_i)\}_{i=1}^{n} \mid y_i \in \{-1, +1\}}$

where D — Data set with 'n' no of Datapoints.



→ In the Diagram line L_2 which is BEST SEPARATE's the data.

→ $mx+c \approx W_x^T + W_0 = 0$. Here we can write the line $mx+c$ as $w^T.x + w_0$

$y = mx + c$

$ax + by + c = 0$

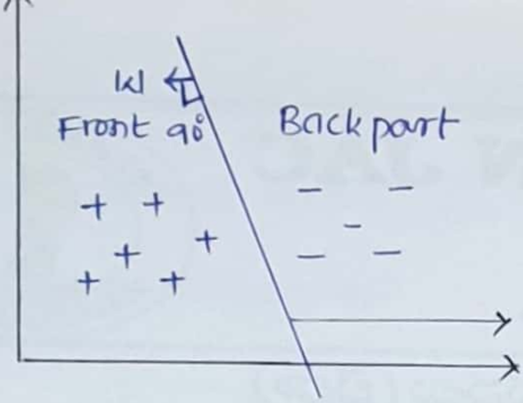
$w_1x_1 + w_2x_2 + w_0 = 0$

$\boxed{w^T.x + w_0 = 0}$

↳ Simply Dot product

* The w is always perpendicular to the line, Role of w is to give the direction of the line where it is facing.

* Based on w Direction is the front part of line and Other Side is backpart of line



w is called of norm of hyperplane | Line, it is nothing but perpendicular line

w is parpendicular to line, which represents the where it is facing (direction)

Here we represent the angle (90°) ? why we choose 'w' is parpendicular

$w^T x + w_0 = 0$

We simply can write as

$[w_1\ w_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + w_0 = 0$

$w_1 x_1 + w_2 x_2 + w_0 = 0$

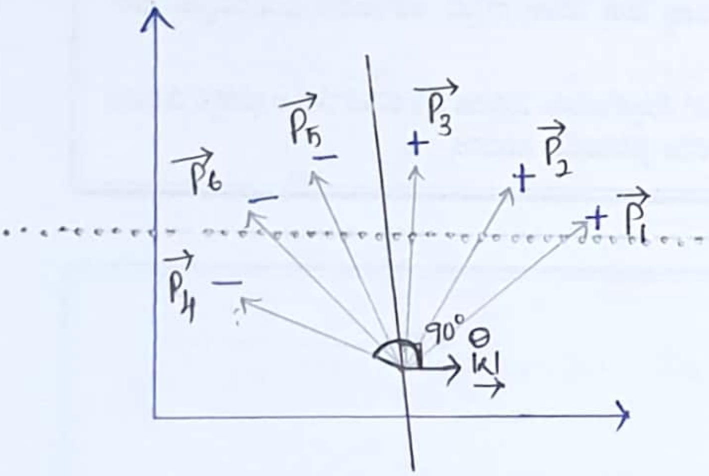
$(w_1 x_1 + w_2 x_2 + w_0) * 1 = 0$

it multiply with any $\in R$

$\Rightarrow \underbrace{[w_1\ w_2\ w_0]\begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}}_{} = \underline{0}$

Nothing but dot product

When ever dot product is equal to '0' (zero) then the angle is 90°



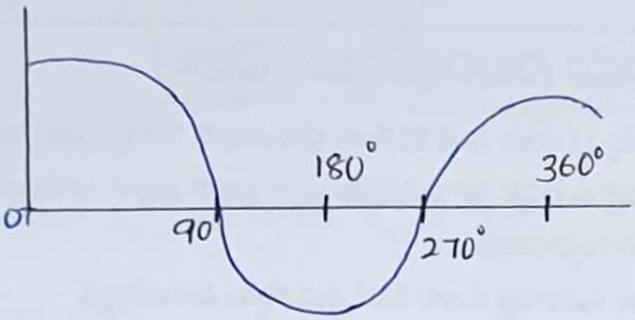
$\Rightarrow$ from the figure –

$\vec{w} \cdot \vec{p} = \underbrace{\|\vec{w}\|\ \|\vec{p}\|}\ \cos\theta_{wp}$

This two values Never be Negative because their magnitude is always positive

depends on angle b/w point to w may vary as +ve or –ve

Here we Consider $\vec{w}$, , Point $\vec{P_1}$ as a vector, There θ is the angle b/w $\vec{w}$ and vectors $\vec{P_1}, \vec{P_2}, \vec{P_3}$. here angle θ is graterthan zero(0) and lessthan 90°



Hence the value for every point is angle b/w 0° to 90° so it gives the positive value.

So $\|\vec{w}\|$ and $\|\vec{P}\|$ always gives positive value

* if we observe the points $\vec{P_4}, \vec{P_5}, \vec{P_6}$ every vector angle grater-than 90° and less than 180° is always gives Negative value.

$$\vec{w}.\vec{P_1} \cos\theta_{w,P_1} = +ve$$

Simply – $\vec{w}\cdot\vec{P_2} = +ve$

$\vec{w}\cdot\vec{P_3} = +ve$

$\|\vec{w}\|.\|\vec{P}\|.\cos\theta \ (0^\circ < \theta < 90^\circ)$ — Here dot product is +ve

$\vec{w}\cdot\vec{P_4} = -ve$

$\vec{w}\cdot\vec{P_5} = -ve$

$\vec{w}\cdot\vec{P_6} = -ve$

$\|\vec{w}\|.\|\vec{P}\| \cos\theta \ (90^\circ < \theta < 180^\circ)$ — Here the dot product is −ve

Dot product gives us the direction of the point wether it is in positive side or in negative side

* Here key point is to Dot product by this we are going to predict the future data point is +ve or −ve ~~side~~ point

if the Dot product is zero the point is lies on the line.

$$y_i * \hat{y}_i = 0 \Rightarrow \text{point lies on line } (y = mx + c)$$

↳ y_{actual} ↳ $y_{predict}$

for $P_1 \to$ y_{actual} is (+ve) * $y_{predict}$ is $w \cdot P_1$ (+)

y_{act} is the label of the point P_1 ie +ve

y_{pred} is value given by the line $(w_x^T + w_0)$, the dot product of w & the point (given) as +ve, for P_1 total value is +ve

for $P_4 \to$ y_{act} (−ve) * y_{pred} $w \cdot P_4$ (−)

y_{act} point (P_4) is the label point which is (−ve)

y_{pred} is dot product of w & the point is also (−ve) then —

Total value is (+ve)

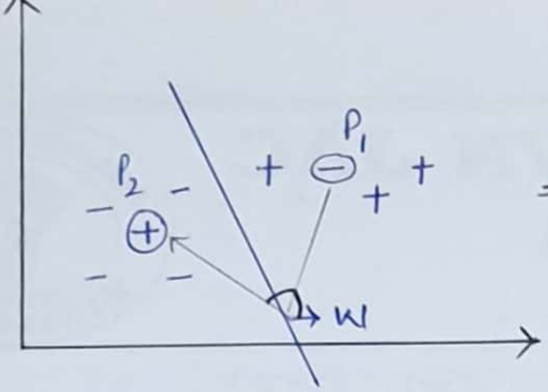
So Similarly do for all points (P_2, P_3, P_5, P_6) → {+, +, −, −}

Hence from this we can say that the act value & pred. values are correctly classified.

* if we take misclassified points ie if there +ve's in Negative region (or) −ve points in positive region the $y_{act} \times y_{pred}$ is always Negative

for correctly classified points $y_{act} \times y_{pred}$ is always positive

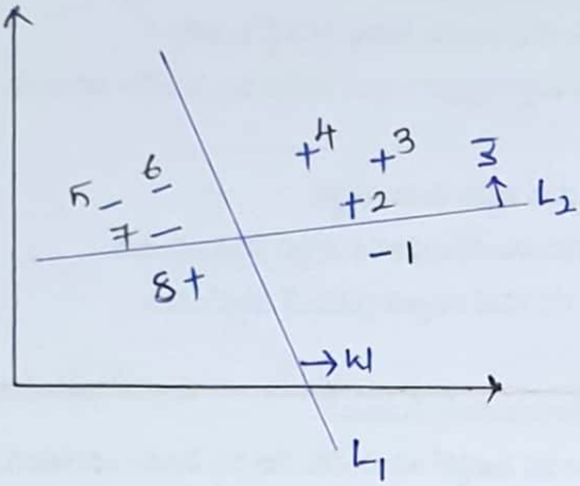
for missclassified points $y_{act} \times y_{pred}$ is always Negative



$\Rightarrow$ $y_{act} * y_{pred}$

for $P_1 \Rightarrow - * w \cdot P_1(+) \Rightarrow$ T. value $= -ve$

for $P_2 \Rightarrow + * w \cdot P_2(-) \Rightarrow$ T. value $= -ve$



Line (L_1)	$[y_{act} \times y_{pred}]$	Line(L_2)
Point 1 $\rightarrow$ $-ve$		$1 \rightarrow +ve$
$2 \rightarrow +ve$		$2 \rightarrow +ve$
$3 \rightarrow +ve$		$3 \rightarrow +ve$
$4 \rightarrow +ve$		$4 \rightarrow +ve$
$5 \rightarrow +ve$		$5 \rightarrow -ve$
$6 \rightarrow +ve$		$6 \rightarrow -ve$
$7 \rightarrow +ve$		$7 \rightarrow -ve$
$8 \rightarrow -ve$		$8 \rightarrow -ve$

L_1 have 2 micclassifications $L_2 \rightarrow 4$ mis. C

if we sum the all the points we get $L_1 \Rightarrow 4$ (+ve values) & $L_2 \Rightarrow 0$ values

from this we can say that we have to maximise the sum of $y_{act} \times y_{pred}$

then we can decrease the missclassifications.

from L_1 line sum value is more hence misclassifications are less

from L_2 line sum of values is zero hence misclassifications are more

Simply $\rightarrow$ $\boxed{\text{Max}\left\{\sum_{i=1}^{n} (y_{act_i} \times y_{pred_i})\right\}}$

So Here our task is to find a Line that BEST SEPARATE +ve from -ve

Line $\downarrow$ m^*, c^*

BEST SEPARATE $\downarrow$ minimize the misclassification (or) Maximuize the Correct classifications.

The m^*, c^* values of a line is Selected in Such a way that optimizes the misclassifications (or) maximize the Correct classifications.

For logistic Regression Eqn Can be written as -

$$\boxed{m^*, c^* = \arg_{m,c} \max \left\{ \sum_{i=1}^{n} \boxed{y_{act(i)} * (mx_i + c)_{pred}} \right\}}$$

Signed Distance

Eqn called as Max Signed distance

Here - +ve - Correctly Classified points

-ve - Misclassified points

from m^*, c^* we can Simply Say that if any future datapoint Come we can easily predict the class.

For linear Regression —

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

↳ Eqn is called as Ordinary least Square (OLS)

Scanned with CamScanner

Linear Regression & Logistic Regression :-

$$\boxed{m^*, c^* = \arg_{m,c} \max \left\{ \sum_{i=1}^{n} (y_i * \underbrace{(mx_i + c)}) \right\}}$$

↓

This getting impact with outliers

also called as Cost function

Max Signed Distance

or

Max Correctly classified points.

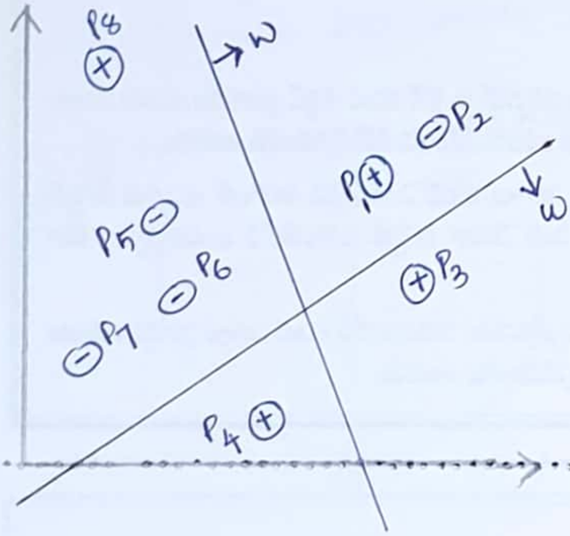
→ The above eqn gives incorrect line when ever outliers are present in the data., model is impacted because of Outlier then the resulting m, c values changed.

$y_{act} \times y_{pred}$ → Signed Distance → it is distance b/w point from the plane

↓

⊕ Correctly classified points

⊖ miss classified points



⇒ We are calculating $y_{act} \times y_{pred}$

	black line	blue line
	2 misclassified	3 misclassified
P_1	→ $\cancel{-1}$	→ $\cancel{+1}$
P_2	→ $\cancel{+1}$	→ $\cancel{-2}$
P_3	→ $+1$	→ $\cancel{+1}$
P_4	→ $+1$	→ $\cancel{-2}$
P_5	→ $+2$	→ $\cancel{+1}$
P_6	→ $+1$	→ $\cancel{+1}$
P_7	→ $+2$	→ $\cancel{+2}$
P_8	→ -10	→ $\cancel{-2}$
	-3	0

→ from Graph we have

black line Contant 2 m.c points

blue line Contain 3 m.c. points

from the Eqn black line has Maximum Signed distance and more misclassified points but {Our aim is to draw a line that has maximum Signed distance and has less misclassified points.} by this the Outliers, the model is getting impacted

Treatment of Outliers in logistic Regression :-

if Outliers, ~~there~~ are there the distance is very large from line or plane

we say that $+\infty$ to $-\infty$ in (max and min outliers)

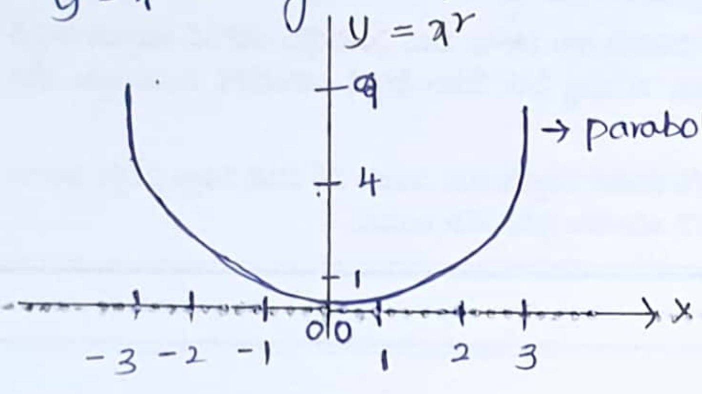
Signed distance is very large $(+\infty)$ or very Small $(-\infty)$ if Outliers are extreme high or extreme low.

Max Σ Signed Distance

goes to $-\infty$ for Extreme low Outlier

goes to $+\infty$ for Extreme high outlier

$y = x^2$ — gives a parabola.



Note — parabola NOT taking care of the Extreme values but it increasing the extreme values.

When $x = 1 \rightarrow y = 1$; $x = -1 \rightarrow y = 1$

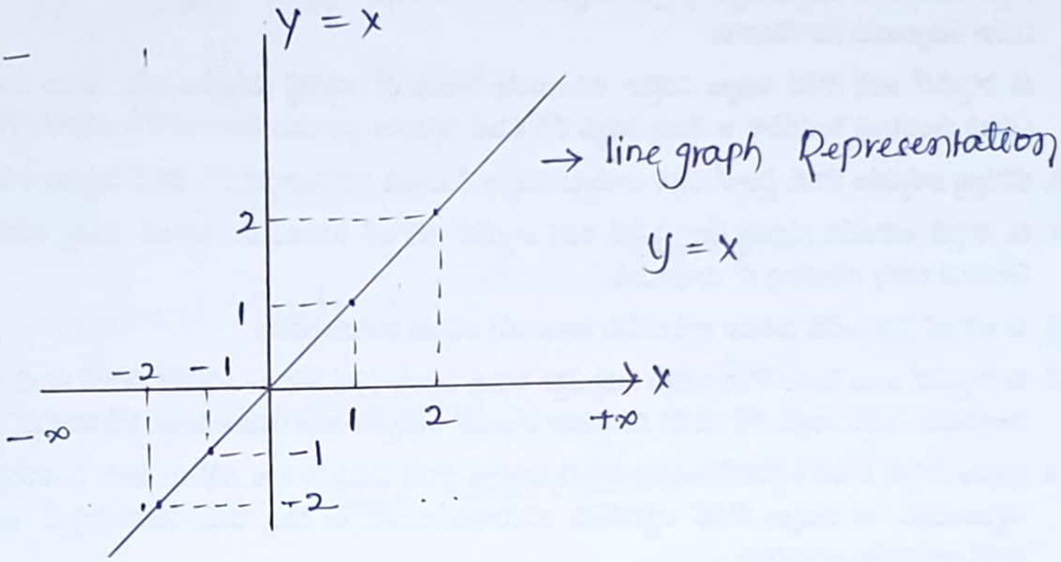
$x = 2 \rightarrow y = 4$; $x = -2 \rightarrow y = 4$

$x = 3 \rightarrow y = 9$; $x = -3 \rightarrow y = 9$

Here we observe - when x value increasing, then y value also increasing rapidly because of squaring, there y will increase to $+\infty$., when x is increase continiously.

when x value increase in negative direction ($x = -1, -2, -3 \ldots$) then y value also increasing rapidly to $(-\infty)$ in negative direction of x plane (or) axis. from this we can say that there is no taking care of extreme values. here extreme values increases. Hence we cant use the parabola.

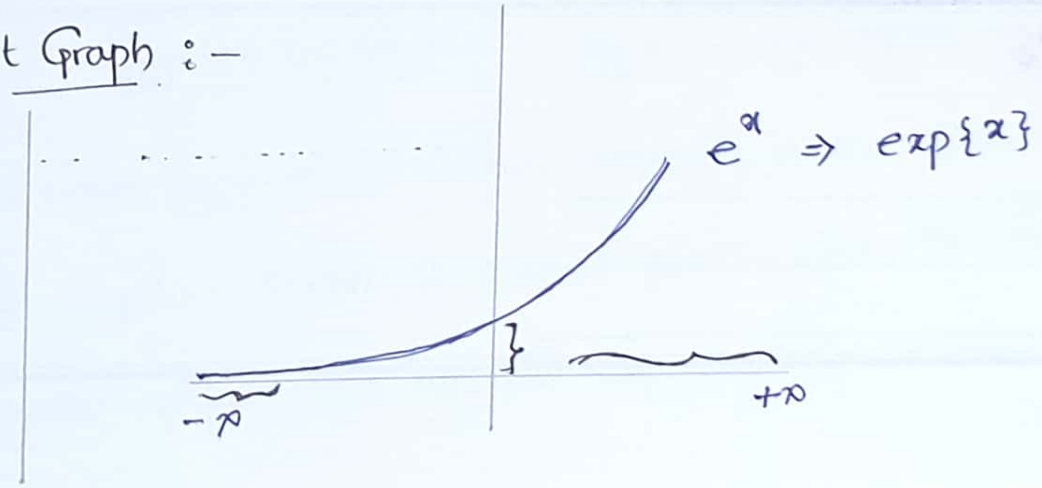
Straight line :-



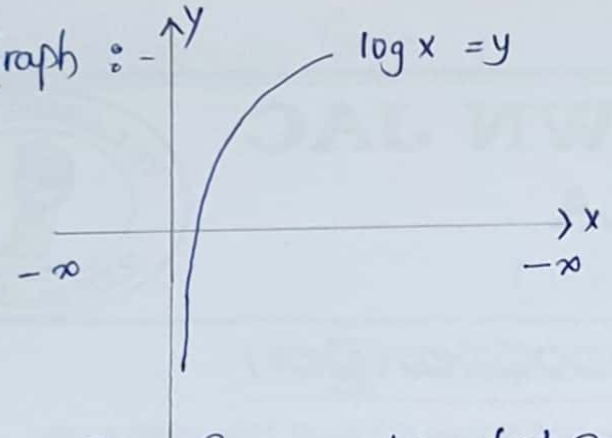
$x = 1 \Rightarrow y = 1$; $x = -1 = y = -1$

$x = 2 \Rightarrow y = 2$; $x = -2 = y = -2$.

Exponent Graph :-



log. Graph :-



Here we Study Some graphs which represent functions of 'y' So here i decide to Use which function (like - $y = x^2$, $y = x$, $y = \log x$ or $y = \exp\{x\}$) apply on Signed distance, So that my Outlier problem is to remove

case - I :- if we apply parabola to function ?

that parabola not taking care of the extreme values and its increase the extreme values So its not Suitable.

Case - II if we apply Exponent Graph ?

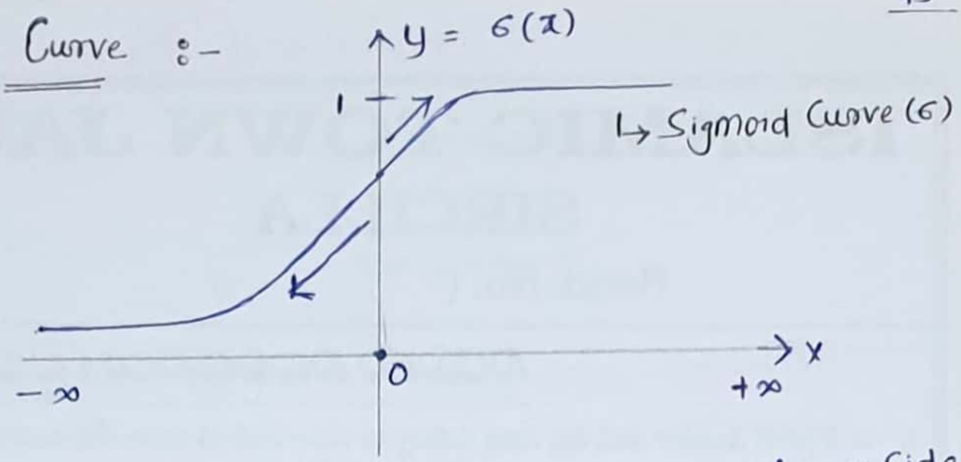
if we apply Exponent Graph ⇒ $e^x = e^{\text{Signed distance}}$

Above Shown graph if we Observe +ve Side of graph Increasing Sharply as Compared to negative side there there is Small increment So in negative Side we can handle by exponent function. , Our problem is on +ve Side So this graph (exponent) has issue On positive Side.

Case - III log function (or) Graph :- log x = log (Signed distance)

if we observe log graph it Contain tappering effect, So and its not defined in negative region, its only On positive Side, for Suppose we have negative values then we Can't represent So its also not a best option.

Sigmoid function (or) Curve :-



if we observe this Curve it contain area of +ve side and -ve side aswell, if +ve region has large no of outliers as compare to -ve side we can handle it by this function (in above three cases we can't treat Outliers) but this Sigmoid Curve we can handle the Outliers.

if we observe the line 'y' Curve rapidly Increases to -ve Side to +ve Side then at Certain point $y_{(max)}$ its Constant Similarly in -ve side also rapidly decreases and Constant at Certain point. ($y_{max} = 1$) ($y_{min} = 0$)

this Sigmoid function denoted by "(σ)" Sigma.

Sigmoid function Can be represent as – $\sigma(x) = \frac{1}{1+e^{-x}}$

from Graph we can Say that the Curve lying always on – {0 to 1}

$$0 \leq \sigma(x) \leq 1$$

↳ Signed Distance

So by this Sigmoid function ~~we can treat Outliers~~ treatment of outlier will be done.

So we have to apply this in logistic regression eqn → we get m^*, c^* value which Outlier treatement after that will be Our problem statement

In eqn we have problem in Signed distance.

Eqn – $\boxed{m^*, c^* = \arg_{m,c} \operatorname{Max} \sum_{i=1}^{n} (y_i * \underbrace{(mx_i + c)})}$

Signed distance → problem statement

Apply Sigmoid-function – $= \dfrac{1}{1+e^{-\text{Signed distance}}}$

e^x – canbe written as – $\exp\{x\} \Rightarrow \dfrac{1}{1+\exp\{-\text{Signed distance}\}}$

Now Eqn will be – $\boxed{m^*, c^* = \arg_{m,c} \operatorname{Max} \sum_{i=1}^{n} \sigma(y_i * (mx_i + c))}$

$$\Rightarrow \operatorname{Max} \left\{ \sum_{i=1}^{n} \frac{1}{1+\exp\{-y_i * (mx_i + c)\}} \right\} \rightarrow \text{Sigmoid function.}$$

By this eqn we can Separate the Class by Suitable line (ie Correctly Classify the class), IB this eqn is more Sueptable to Outliers.

In real world if the data is like this :–

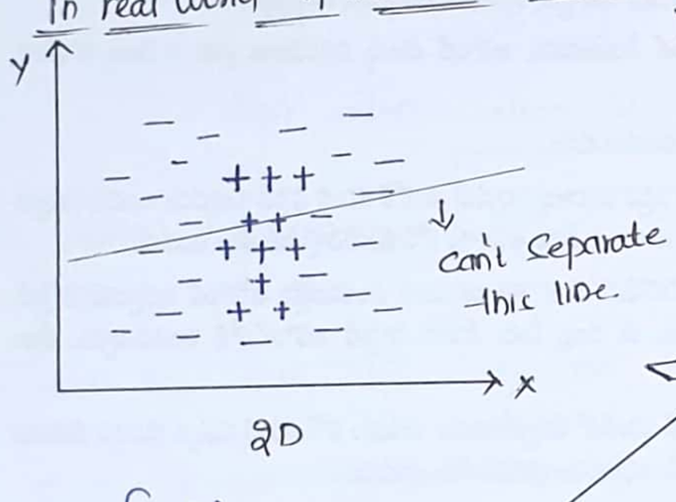
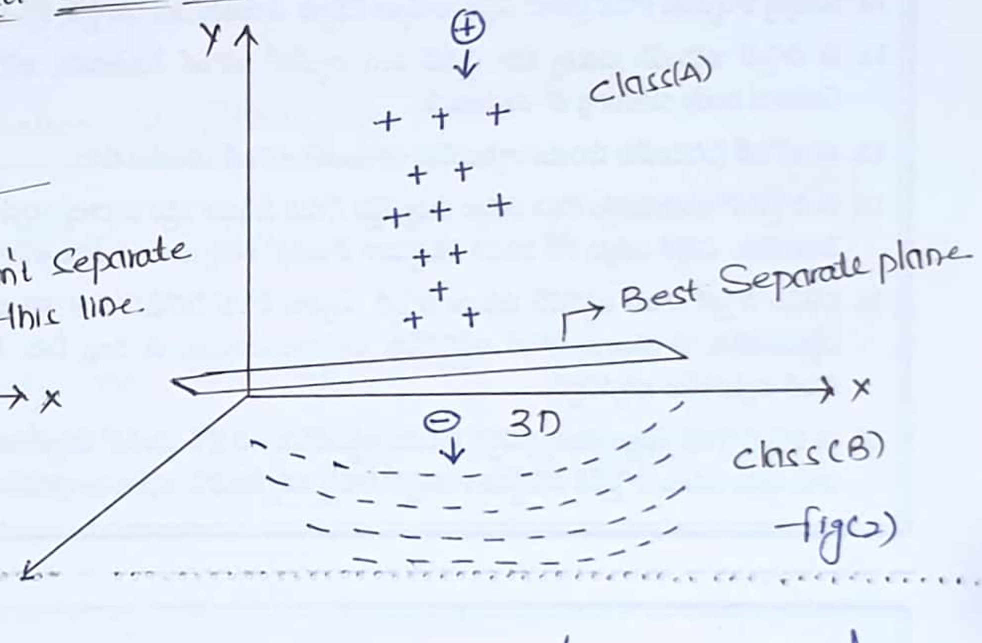


fig (1)



fig(2)

If we observe fig-1 here the two class point's are arranged like Overlapping on One Class if we have to draw any line On that that is not Separated exactly in two class and that line is not Separate in best way Same problem if we have in 3-D Senorio.

the diagram is look like fig(c) here that plane easily separated class by two class.

So finally here we optimizing the m^*, c^* value in eqn ie Maximize the Sum of Sigmoid Signed distance.

Note: In optimization theory, there are some values, if we are trying to maximize the function it is nothing but equal to minimizing the increase of that function.

$$\boxed{\Rightarrow \text{Max } f = \min \frac{1}{f} \quad \Rightarrow \max f = \min -f}$$

⇒ Minimizing the logistic Regression

$$\boxed{m^*, c^* = \arg_{m,c} \sum_{i=1}^{n} 1 + \exp\{-y_i * (mx_i + c)\}}$$

↳ This is Optimizing theorey by this we are minimize function

↳ Eqn also called as Optimizing Eqn.

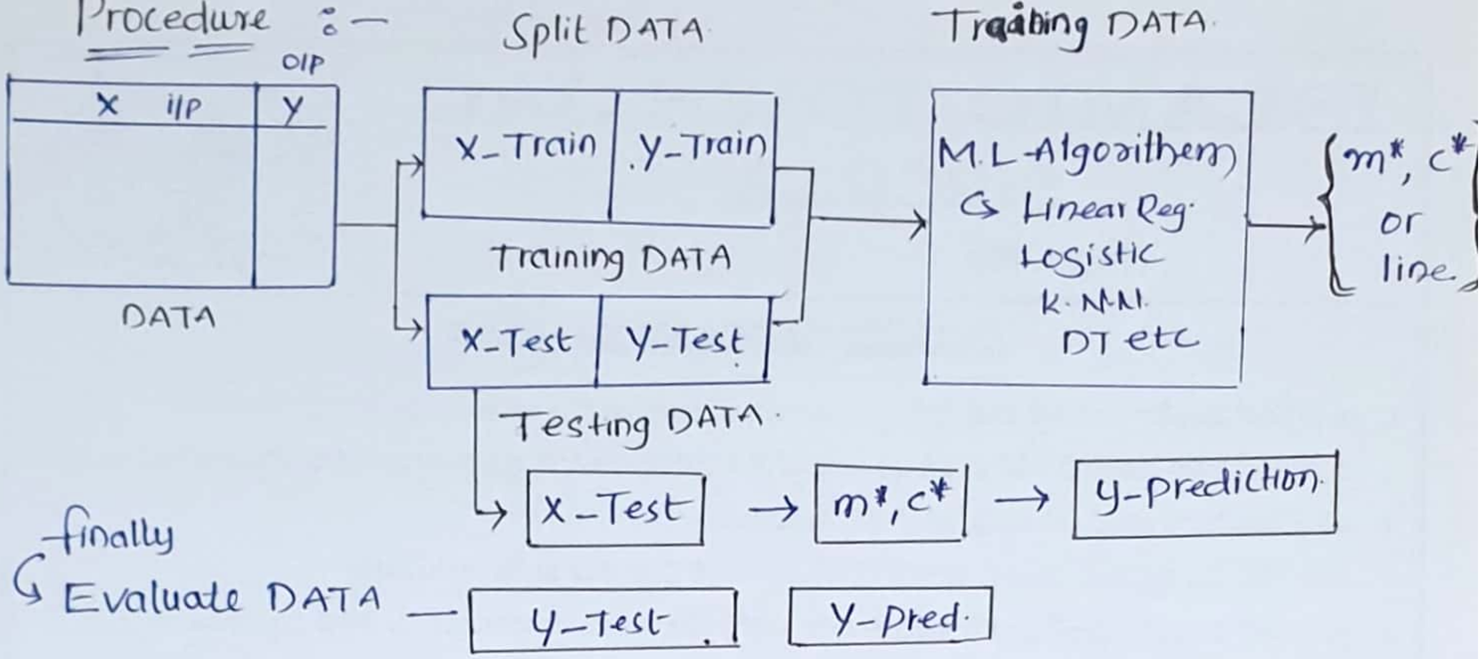
In linear Regression – $\boxed{m^*, c^* = \arg_{m,c} \text{Min}\{\Sigma (y_i - (mx_i + c))^2\}}$

So by this two eqn's we can optimizing (min, max values)

⇒ ie minimizing the Sum of Squared Error → linear Regression

& Maximize the Sum of Sigmoid (Signed distance) – logistic Reg.

Procedure :-

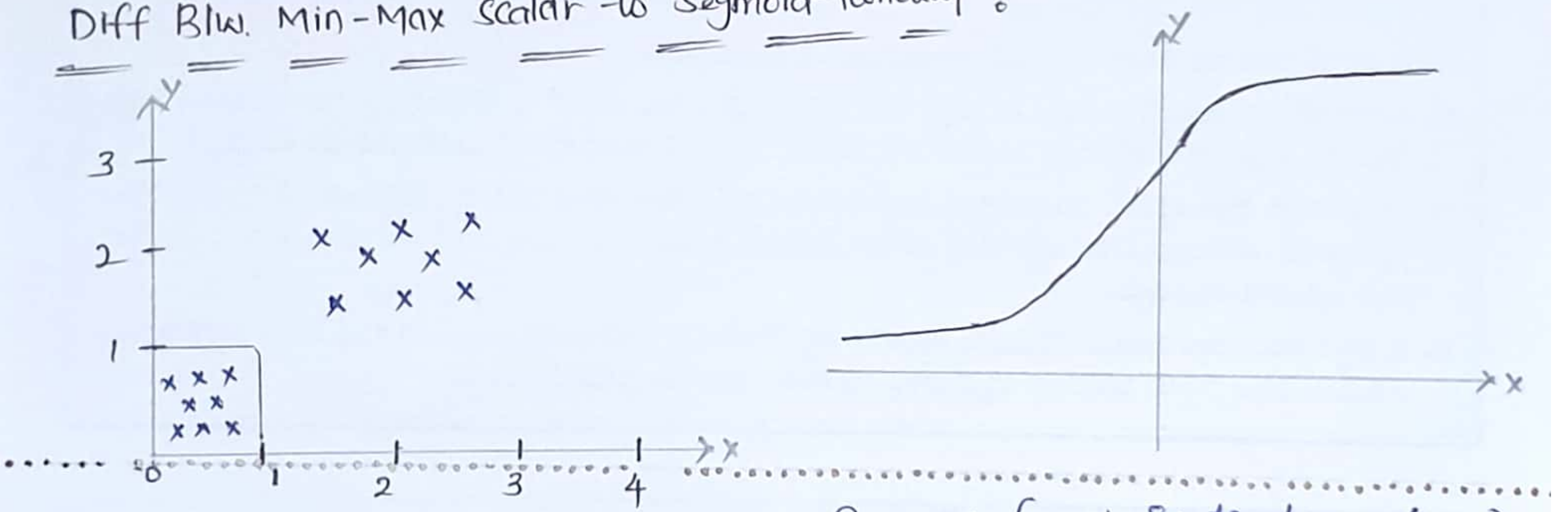


* KNN | Linear Regression | Logistic Regression

Used for classification task → Regression Task → classification Task

finding K-Nearest Neighbors → Best fit line → best separate classes

Diff B/w. Min-Max scalar to Segmoid function :-



Key diff b/w (min-max) Scalar to Segmoid funct is to In (m-m) Scalar the whole data Shrink and fit into Scale of (0,1),

In Segmoid basically it won't effect or Impact to the non-outlier points but if we get Outlier as soon we move +ve Side, -ve Side extreme values it is just try to lapping it. with slowly as shown above