MACHINE LEARNING DAY 4 16/10/2022

- Logistic Regression (classification Problem)

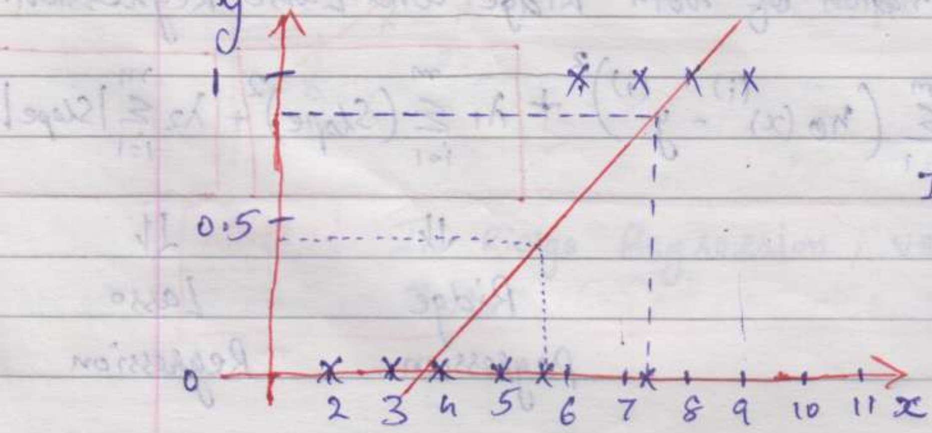
~~-~~ It is used to solve the classification problem

ITT JEE	Study Hrs	Play Hrs	o/p (Pass/Fail)
	1	8	Fail
	2	7	Fail
	3	7	Fail
	6	3	Pass
	1	4	Pass (outliers)

DATASET

Study Hours	O/P (Pass/Fail)
2	FAIL (0)
3	FAIL (0)
4	FAIL (0)
5	FAIL (0)
6	PASS (1)
7	PASS (1)
8	PASS (1)
9	FAIL (1)

Que Can we solve this problem using Regression?



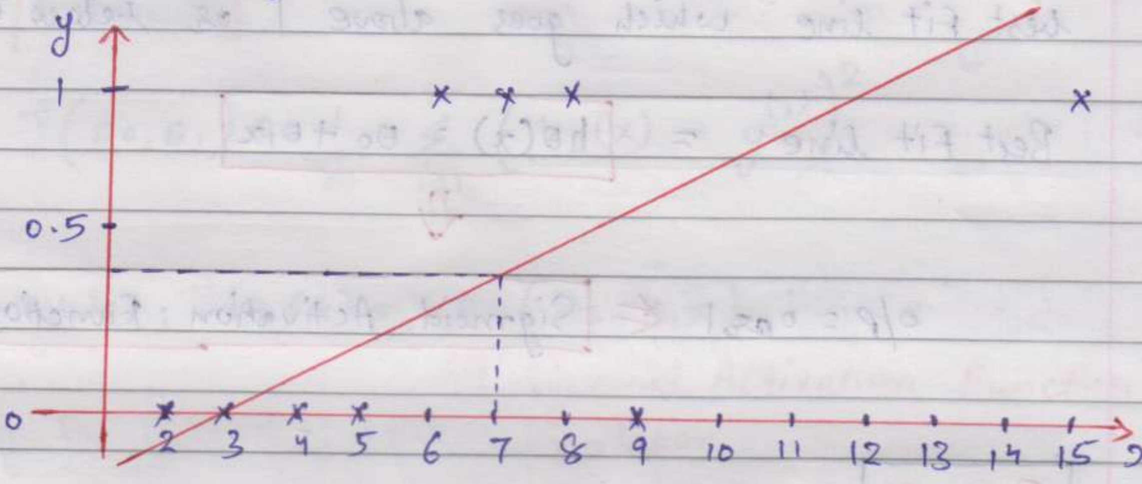
0.5 => Threshold.

If $y \leq 0.5 = 0$

$y > 0.5 = 1$

Que: Why we cannot use Linear Regression here?

- Suppose in the dataset one Person study for 15 hours and he will pass then our best fit line will change



- So Even if a person study for 7 Hours $y < 0.50$ therefore the model will show Fail.
That is why we cannot used Linear Regression in this type of problem statement.

Note :-

Logistic Regression: is applied to predict the categorical dependent variable. In other words, it's used when the prediction is categorical, for example, yes or no, true or false, 0 or 1. The predicted probability or output of logistic regression can be either one of them, and there's no middle ground.

Que: Difference between Logistic Regression Vs Linear Regression?

Ans:- Linear Regression is used to predict the continuous dependent variable using a given set of independent variables.

Logistic Regression is used to predict the categorical dependent variable using a given set of independent variables.

Sigmoid Activation Function is used to squash the best Fit line which goes above 1 or below 0

Best Fit line = $\boxed{h\theta(x) = \theta_0 + \theta_1 x}$

$\Downarrow$

o/p = 0 or 1 $\Leftarrow$ $\boxed{\text{Sigmoid Activation Function}}$

Formula

$$\boxed{\text{Sigmoid Function} = \frac{1}{1 + e^{-Z}}} \left\{ Z = \theta_0 + \theta_1 x \right\}$$

Linear Regression Cost Function

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

$h\theta(x) = \theta_0 + \theta_1 x$

$\boxed{\text{MSE}}$

$\Downarrow$

$\boxed{\text{Convex Function}}$

$\Downarrow$

1 Global Minima

Logistic Regression Cost Function

① Create a Best Fit Line

② Squassing using Sigmoid Activation Function

Formula

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

where : $h_\theta(x) = \sigma(\theta_0 + \theta_1 x)$

↳ Sigmoid Activation Function notation

Suppose $Z = \theta_0 + \theta_1 x$

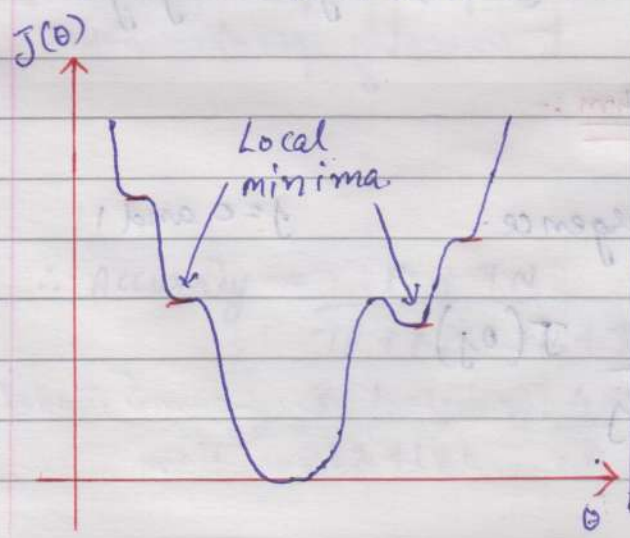
Sigmoid Activation Function $(\sigma) = \frac{1}{1+e^{-Z}}$

$$\therefore h_\theta(x) = \sigma(z)$$

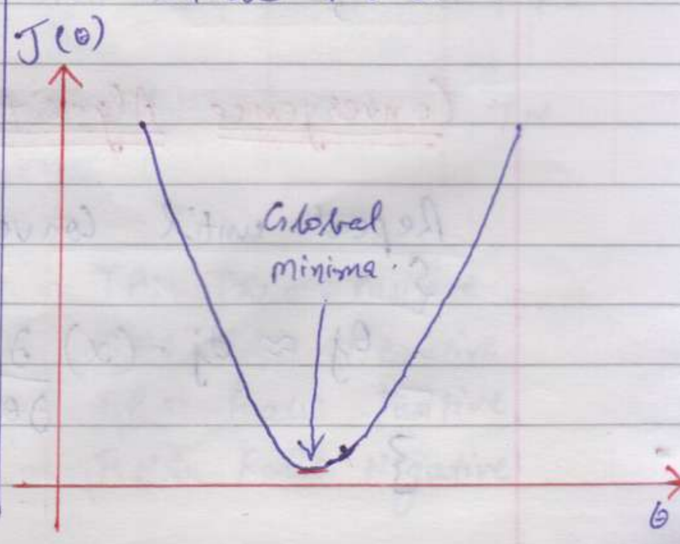
$$= \frac{1}{1+e^{-Z}}$$

$$\therefore \boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}}$$

Non - Convex Function



Convex Function



- The problem with Cost function of Logistic Regression is that we get result like Non - Convex function

- The simple way to fix this problem is to simply change the cost function.

★ Log Loss Cost Function

$$\text{Cost}\left(h_\theta(x)^{(i)}, y^{(i)}\right) = \begin{cases} -\log\left(h_\theta(x)\right) : \text{if } y = 1 \\ -\log\left(1 - h_\theta(x)\right) \text{ if } y = 0 \end{cases}$$

$$\boxed{\text{Cost}\left(h_\theta(x)^{(i)}, y^{(i)}\right) = -y\log\left(h_\theta(x)\right) - (1-y)\log\left(1-h_\theta(x)\right)}$$

$$\Downarrow$$

This Cost Function will create a convex function and we never get a local Minima.

$\boxed{y : \text{Truth value}}$

Minimize Cost Function $J(\theta_0, \theta_1)$, by changing θ_0, θ_1

Convergence Algorithm :-

Repeat until Convergence. $j = 0$ and 1

$$\{$$

$$\theta_j \approx \theta_j - (\alpha)\frac{\partial}{\partial\theta_j} J(\theta_j)$$

$$\}$$

Performance Metrics :-

① Confusion Matrix
② Accuracy
③ Precision
④ Recall
⑤ F-Beta Score

DATASET

f1	f2	o/p (y)	ŷ (Model Prediction)
-	-	0	1
-	-	1	1
-	-	0	0
-	-	1	1
-	-	1	1
-	-	0	1
-	-	1	0

=> Confusion Matrix

(y) Actual value

$$\begin{array}{cc|c|c|} & & 1 & 0 \\ \hline (\hat{y}) \text{ predicted value} & 1 & 3 & 2 \\ \hline & 0 & 1 & 1 \\ \hline \end{array}$$

(y) Actual value

$$\begin{array}{cc|c|c|} & & 1 & 0 \\ \hline \hat{y} \text{ Predicted value} & 1 & TP & FP \\ \hline & 0 & FN & TN \\ \hline \end{array}$$

$$\therefore \text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$= \frac{3+1}{3+2+1+1} = \frac{4}{7}$$

TP: True Positive
TN: True Negative
FP: False Positive
FN: False Negative

∴ Accuracy = 57%

* DATASET → Binary Classification

Lets say we have 1000 datapoints

$$1000\ \text{datapoints}\begin{cases}\rightarrow 900 \rightarrow 1\\ \rightarrow 100 \rightarrow 0\end{cases}\Bigg\}\ \text{Imbalanced Dataset}$$

Now If we make any Dumb model that predict 1 all the time.

$$\text{Dumb Model} \rightarrow 1 \ \Big\}\ 90\% = \text{Accuracy.}$$

So we get 90% Accuracy, which is not good. Therefore we cannot depend only on accuracy.

⊛ Precision :-

$$\boxed{\frac{TP}{TP+FP}}$$

Actual value

	1	0
1	TP	FP
0	FN	TN

Predicted value

Question:- Out of all the actual values how many are correctly predicted.

Problem Statement:

Received mail is a Spam or Ham.?

TP: Mail is Spam and Predicted as Spam { Model is Good }

FN: Mail is Spam but Predicted as Ham

{Not Good but it is OK}

FP: Mail is not Spam but predicted as Spam {Cretical Problem}

TN: mail is Ham and predicted as Ham.

⇒ Therefore:- Our main focus should be ~~so~~ on reducing FP {False Positive}

⇒ But if we have to make model to predict a person is having disease or not, we have to focuss on reducing FN {False Negative}

(*) **Recall** :- $\boxed{\frac{TP}{TP + FN}}$ ⇒ Out of all the predicted values how many are correctly Predicted

Suppose, Tomorrow the stock market is going to crash What we have to focus on FP or FN ?

	1	0
1	TP	FP
0	FN	TN

→ Consumer → FN↓↓

→ Companies → FP↓↓

⊛ F - Beta Score :-

$$\boxed{F_\beta = \frac{(1+\beta^2)(\text{precision} * \text{recall})}{(\beta^2 \cdot \text{precision} + \text{recall})}}$$

Condition 1

If FP and FN, both are important

then $\beta = 1$

$$F_1 \text{ Score} = \frac{(2)(\text{precision} * \text{recall})}{\text{precision} + \text{recall}}$$

Condition 2

If FP is more important then FN

then $\beta = 0.5$

$$F_{0.5} \text{ Score} = \frac{(1+0.25)(\text{Precision} * \text{recall})}{(0.25 \cdot \text{precision} + \text{recall})}$$

Condition 3

If FN is more important then FP

then $\beta = 2$

$$F_2 \text{ Score} = \frac{(1+4)(\text{Precision} * \text{recall})}{(4 * \text{precision} + \text{recall})}$$