ML-DAY-4

LOGISTIC REGRESSION

(Classification Problem)

Let we have a dataset as follows.

Students will take IIT/JEE exam.

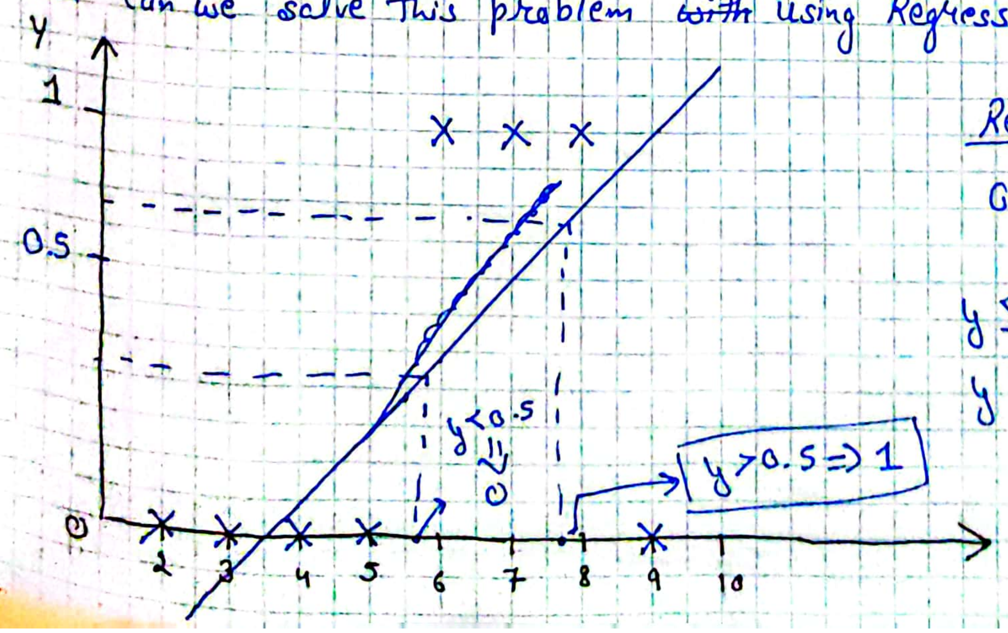
	Study	Play Hrs	O/P (Pass/Fail)
	1	8	Fail
	2	7	Fail
	3	7	Fail
	6	3	Pass
Outlier ←	1	4	Pass

Make the machine predict the output based on independent features.

→ Eg:- Dataset UPSC

Study Hrs	O/P (Pass/Fail)
2	Fail
3	Fail
4	Fail
5	Fail
6	Pass
7	Pass
8	Pass
9	Fail

Q Can we solve this problem with using Regression?



Regression

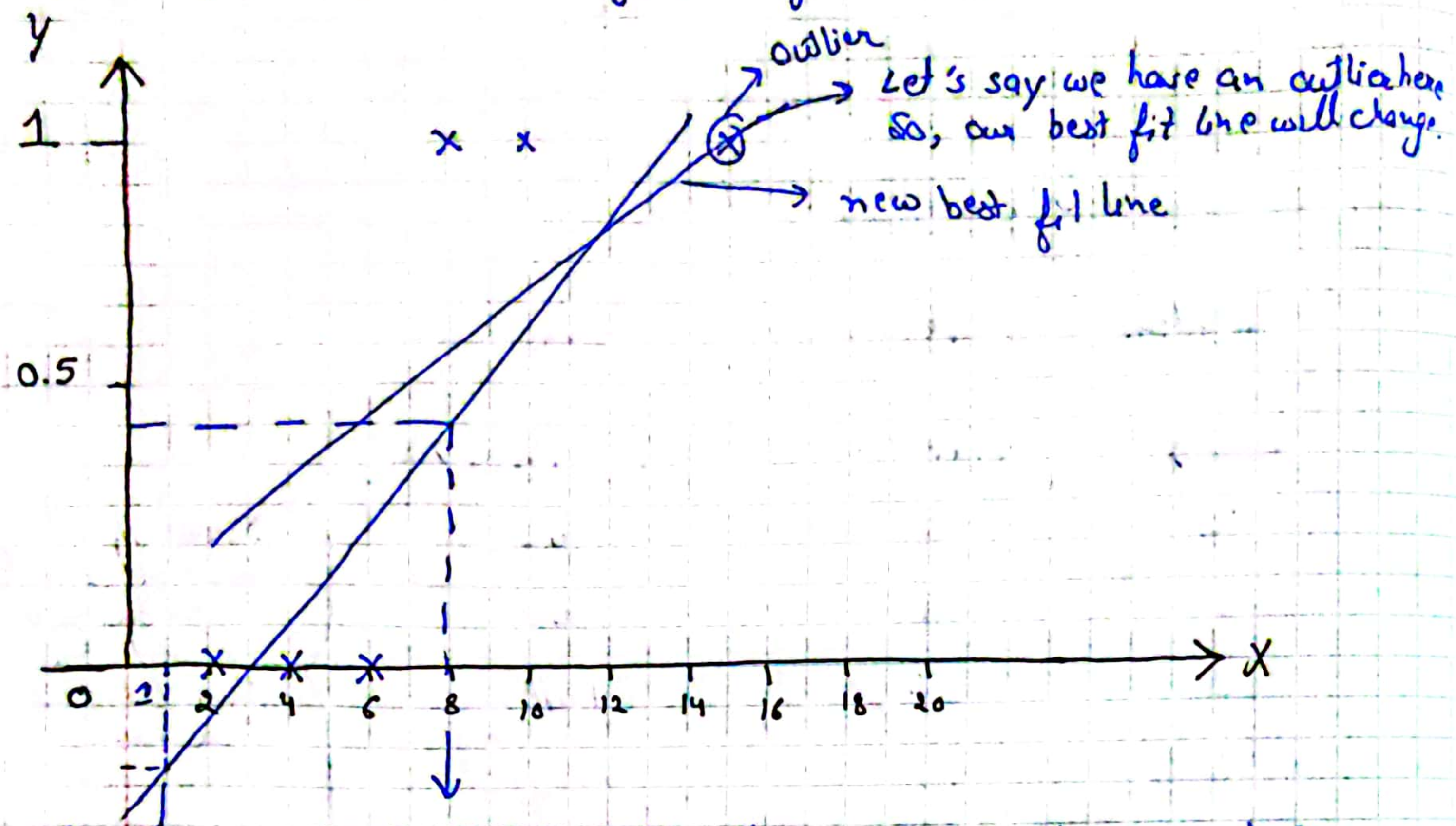
0.5 ⇒ Threshold

$y \leq 0.5 = 0$

$y > 0.5 = 1$

Q If we can solve it with ~~the~~ Linear Regression,

* why do we need Logistic Regression?



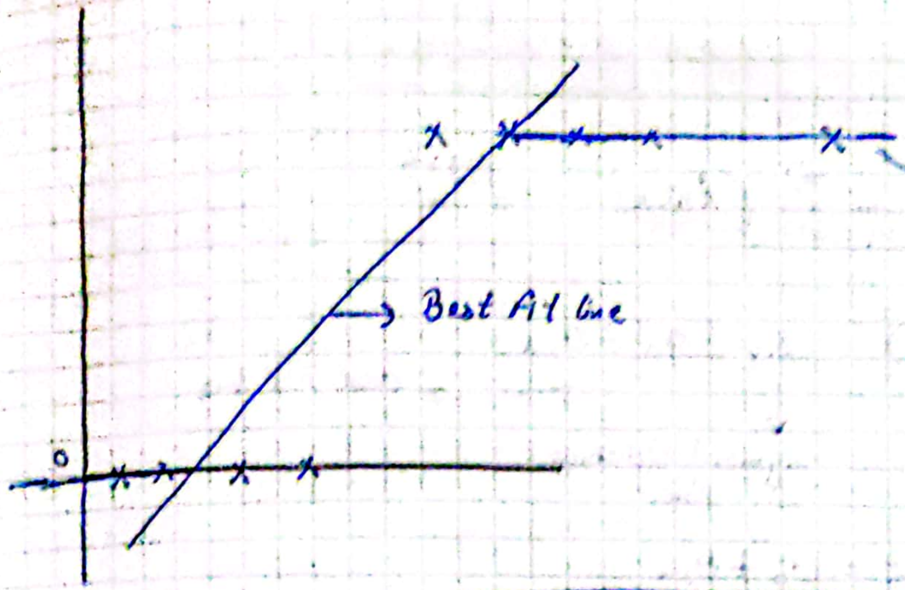
Here Output is less than 0

Here the model will say, even if the student is studying for 8 hrs, he/she will fail.

⇓

* Due to some of outliers my best fit line will completely change which will change the results.

→ Logistic Regression will squash the line b/w 0 and 1 hence used for classification.



Squashing of line

↓

Squashing is done by $\boxed{\text{Sigmoid Activation func}}$

best fit line $= \boxed{h_\theta(x) = \theta_0 + \theta_1 x}$

⇓

Sigmoid Activation ⇒ output = 0 to 1

① $z = h_\theta(x) = \theta_0 + \theta_1 x$

$z = \theta_0 + \theta_1 x$

② Sigmoid function $= \boxed{\dfrac{1}{1+e^{-z}}} \Rightarrow$ 0 to 1

Linear Regression Cost Function

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

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

⇓

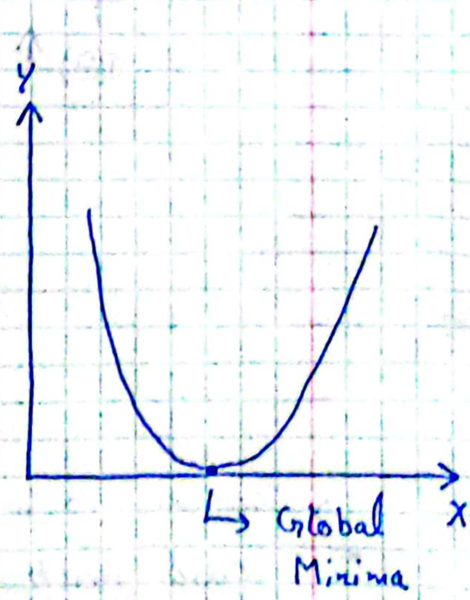
MSE

⇓

Convex function

⇓

1 Global Minima



Logistic Regression Cost Function

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

Here $h_\theta(x) = Sig(\theta_0 + \theta_1 x)$

$\theta_0 + \theta_1 x \rightarrow$ Best fit line

$Sig \rightarrow$ Sigmoid Advance fnc.

Let $z = \theta_0 + \theta_1 x$

$$h_\theta(x) = sig(z)$$

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

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

$\Downarrow$

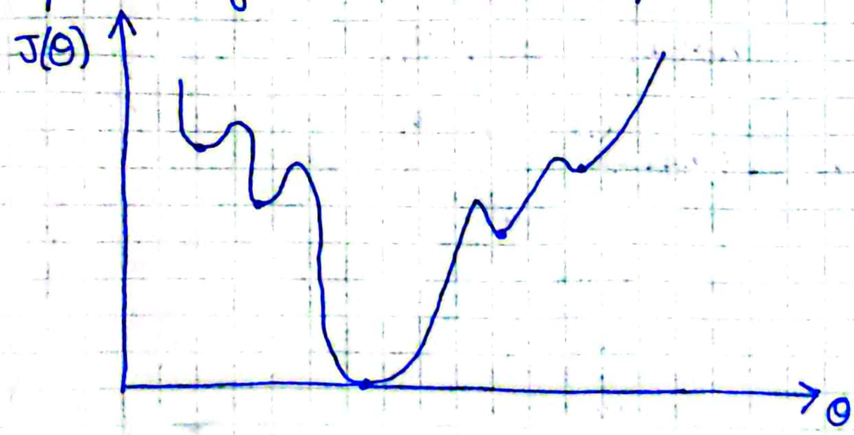
0 to 1

$\Downarrow$

~~There~~ There is a limitation with this equation

$\Downarrow$

This eqn. will give a non-convex function.



$\rightarrow$ The way to solve this problem is to change the

* cost function.

Log Loss Cost Function

$$\text{Cost}(h_\theta(x)^{(i)} - y^{(i)}) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) \end{cases}$$

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

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

if $y = 1$

$$\text{Cost}(h_\theta(x)^{(i)} - y^{(i)}) = -\cancel{y}\log(h_\theta(x)) - 0 = -\cancel{y}\log(h_\theta(x))$$

if $y = 0$

$$\text{Cost}(h_\theta(x)^{(i)} - y^{(i)}) = 0 - (1-0)\log(1-h_\theta(x)) = -\log(1-h_\theta(x))$$

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

$\Downarrow$

Convex Function

$\Downarrow$

Never Local Minima

→ Aim: Minimize Cost fnc $J(\theta_0, \theta_1)$ by changing θ_0, θ_1

$\Downarrow$

Convergence algorithm

Repeat Convergence

$\{$

$j = 0$ and 1

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

$\}$

Performance Metrics

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

→ Confusion Matrix

Dataset

Feature1	Feature2	O/P	$\hat{y}$ → Model Prediction
–	–	0	1
–	–	1	1
–	–	0	0
–	–	1	1
–	–	1	1
–	–	0	1
–	–	1	0

Actual

		1	0
Predicted	1	TP	FP
	0	FN	TN

Here,

Actual

		1	0
Predicted	1	3	2
	0	1	1

→ Accuracy

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

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

$$\boxed{\text{Acc.} \approx 57\%}$$

→ Precision

Let we have a dataset

$$\text{Dataset} \rightarrow \left.\begin{matrix} 900 \rightarrow 1 \\ 100 \rightarrow 0 \end{matrix}\right\} \text{Imbalanced Dataset}$$

If we create a Dumb Model → 1 → 90% accuracy

⇓

Here only accuracy is not sufficient

⇓

We have to focus on Precision, Recall and F-beta score.

* $$\boxed{\text{Precision} = \frac{TP}{TP+FP}} \Rightarrow$$ Out of all the actual values how many are correctly predicted.

→ In precision, we focus on FP and we try to reduce FP

→ Recall

$$\boxed{\text{Recall} = \frac{TP}{TP+FN}}$$

⇒ Out of all the predicted values how many are correctly predicted.

→ F- Beta Score

Eg :- Tomorrow the stock market is going to crash.

⇓

→ Consumer's Perespective → Try to Reduce FN
FN↓↓↓

→ Company's perspective → Try to Reduce FP
FP↓↓↓

F-Beta Score :

$$\boxed{\frac{(1+\beta^2)\,(\text{Precision} * \text{Recall})}{\beta^2 * (\text{Precision} + \text{Recall})}}$$

① If FP and FN ~~are~~ both are important

⇓

$\beta = 1$

$$\boxed{F1\ \text{score} = 2\,\frac{P*R}{P+R}}$$

② If FP is more important than FN

$\beta = 0.5$ ⇒ ~~f1 score =~~

⇓

$$F\,0.5\ \text{score} = \frac{(1+0.25)\,(P*R)}{(0.25)\,(P+R)}$$

③ If FN is more important than FP

$$\beta = 2$$

$$\boxed{F2 \text{ score} = \frac{(1+4)\ (P * R)}{4\,(P+R)}}$$