Notes

Machine learning

Logistic Regression [Classification problem]

Example: IIT / JEE

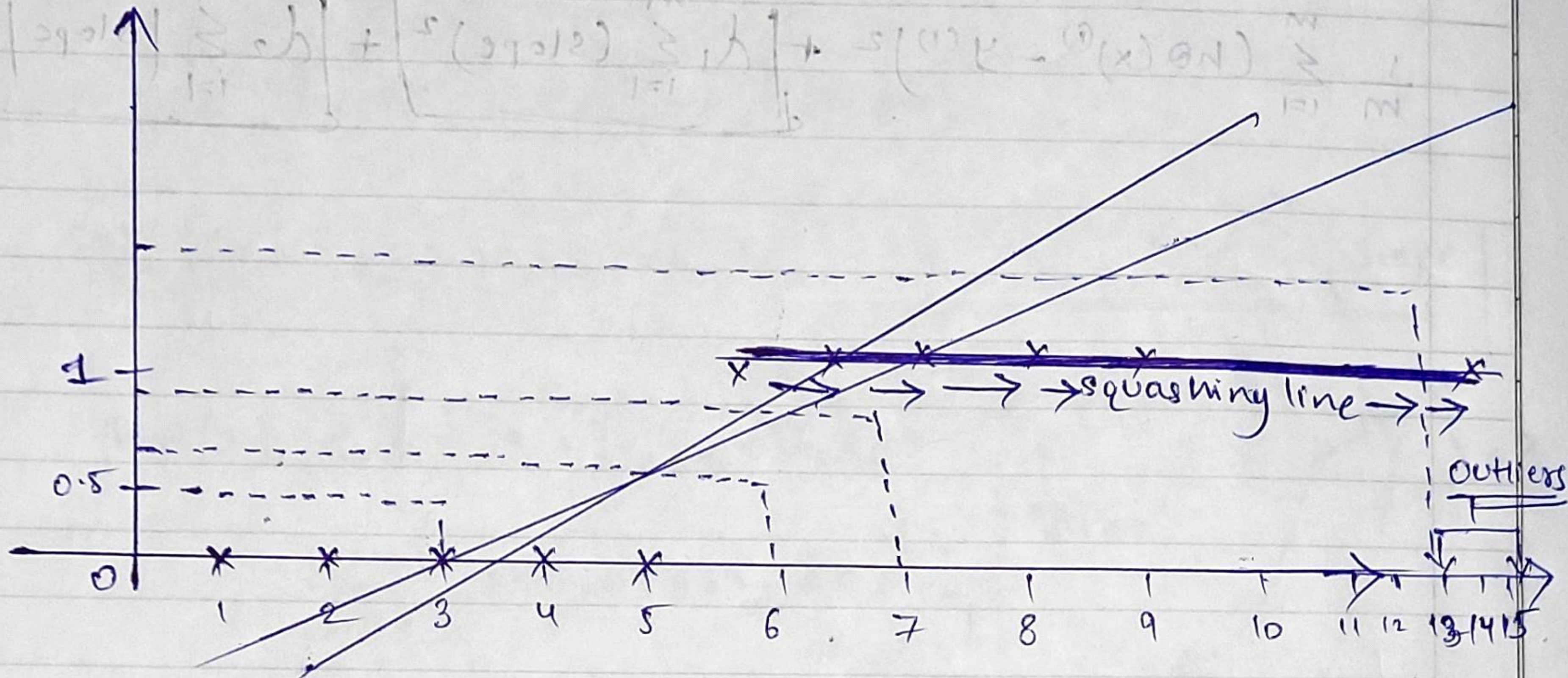
study (hour)	playing (hour)	O/P (Pass/fail)
1	8	fail
2	7	fail
3	7	fail
6	3	Pass
	-	

* No. of study hour increases, chance of Pass will be increase

* There will be some person, who study less but they can also clear Exam, that person consider as outliers.

Ex: DATASET (UPSC)

study (hours)	O/P (Pass/fail)
2	FAIL
3	FAIL
4	FAIL
5	FAIL
6	PASS
7	PASS
8	PASS
9	PASS



conditions :- $y \leq 5$ and $y > 0.5$ (if more than one because of outlier than we use squashing line function [sigmoid Activation]

⇓ 0 ⇓ 1

⇓ Logistic Regression.

why use logistic instead of linear
→ outliers causes changes Best fit line and we have only two condition 0 to 1, it go above 1 then we use sigmoid function which squash line.

Date..............................

Expt. No......................

Page. No......................

Sigmoid Activation:

$\Downarrow$

$\boxed{h_\theta(x) = \theta_0 + \theta_1 x} \Rightarrow o/p = \boxed{0 \text{ to } 1}$

1. $z = h_\theta(x) = \theta_0 + \theta_1 x$ Let $\rightarrow \boxed{z = \theta_0 + \theta_1 x}$
2. sigmoid function $\Rightarrow \boxed{\frac{1}{1+e^{-z}}} \Rightarrow$ 0 to 1

Steps for logistic Regression:

1. Create a Best fit line.
2. squashing $\rightarrow$ Sigmoid function

Linear Regression cost fn	Logistic Regression Cost fn:
$J(\theta_0, \theta_1) = \frac{1}{n} \sum_{i=1}^{n} (h_\theta(x)^{(i)} - y^{(i)})^2$	$J(\theta_0, \theta_1) = \frac{1}{n} \sum_{i=1}^{n} (h_\theta(x) - y^{(i)})^2$
$h_\theta(x) = \theta_0 + \theta_1 x$	$h_\theta(x) = sig(\theta_0 + \theta_1 x)$
$\Downarrow$ MSE $\Downarrow$ convex function	Best fit line $\boxed{z = \theta_0 + \theta_1 x}$
* Gradient Descent	$\boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_0 + \theta_1)}}}$
* 1 Global Minimal	$\boxed{h_\theta(x) = \frac{1}{1+e^{-z}}}$

Teacher's Signature..............................

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

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

$z = \theta_0 + \theta_1 x$

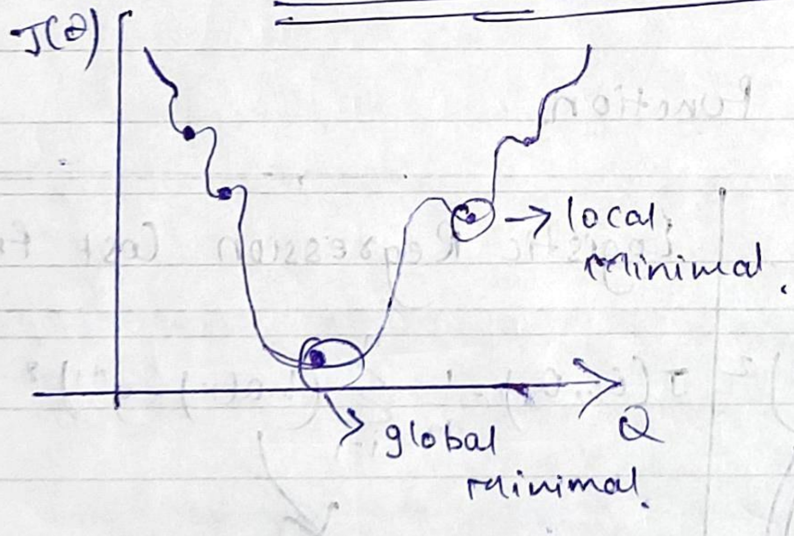
$$\boxed{h_\theta(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}} \Longrightarrow 0 \text{ to } 1$$

$\le 0.5 \Rightarrow 0 \Rightarrow$ Fail

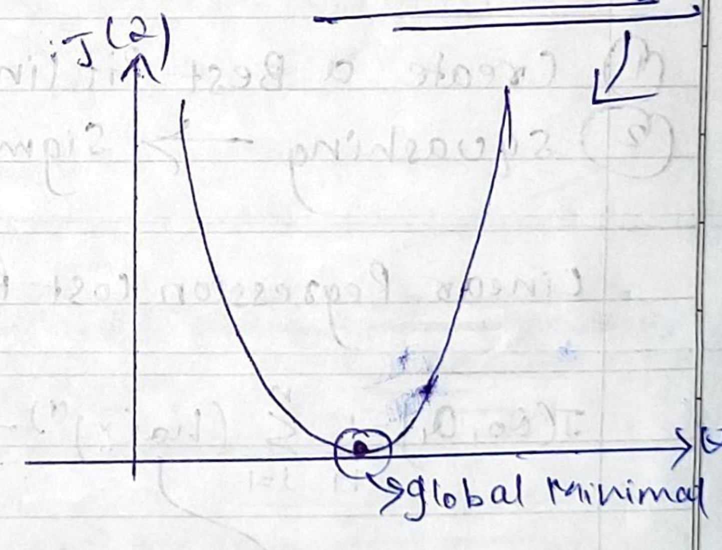
$> 0.5 \Rightarrow 1 \Rightarrow$ Pass

$\boxed{\text{Threshold} = 0.5}$

Non-convex function



Convex function



Expt. No.................... Page. No....................

* Log Loss function:

$$\text{cost}(h_\theta(x)^{(i)}, y^{(i)}) = \begin{cases} \boxed{-\log(h_\theta(x))} & \text{if } y=1 \\ \boxed{-\log(1-h_\theta(x))} & \text{if } y=0 \end{cases}$$

convex function

$\Downarrow$

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

$\Downarrow$

Never local minimal

Minimize cost function $J(\theta_0, \theta_1)$ by changing θ_0, θ_1

Convergence Algorithm.

Repeat convergence

$\{$

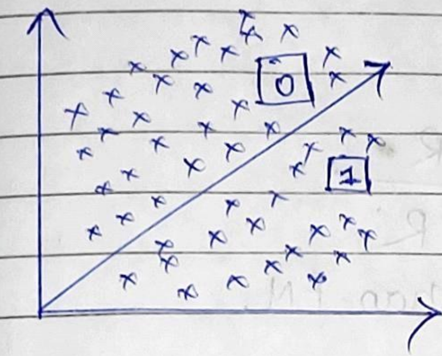
$j = 0$ and 1

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

$\}$

Teacher's Signature....................

Performance Matrix :-



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

DATASET :-

Feature 1	Feature 2	y. O/P Actual	$\hat{y}$ Predicted.
—	—	0	1
—	—	1	1
—	—	0	0
—	—	1	1
—	—	1	1
—	—	0	1
—	—	—	—

Predicted \ y Actual	1	0
1	TP	FP
0	FN	TN

Predicted \ Actual	1	0
1	3	2
0	1	1

⇓

Confusion Matrix

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

Accuracy : $\dfrac{TP + TN}{TP + FP + FN + TN}$

$= \dfrac{3+1}{3+2+1+1} = \dfrac{4}{7} \approx \boxed{57\%}$

* DATASET → BINARY CLASSIFICATION

↳ 1000 Datapoints → 900 → 1, → 100 → 0 $\Big\}$ imbalanced data

Dumb Model → 1 ⇒ 90% Accuracy → ✗ sufficient

* **Precision** : $\boxed{\dfrac{TP}{TP + FP}}$

	Actual 1	Actual 0
Predicted 1	TP	FP
Predicted 0	FN	TN

out of all the actual values how many are correctly predicted.

Model → Diabetes or not Diabetes.

* Recall : $\boxed{\dfrac{TP}{TP + FN}}$ ⇒ out of all predicted values how many are correctly predicted

* Tomorrow, the stock market is going to crash

	1	0
1	TP	FP
0	FN	TN

↓

→ consumer → FN

→ companies → FP

F-Beta score : $\dfrac{1+\beta^2 \quad \text{Presession} * \text{Recall}}{(\beta^2 * \text{Precision} + \text{Recall})}$

(1) If FP and FN are both important

$$\beta = 1$$

$$F1\ Score = 2\frac{P * R}{P + R}$$

(2) If FP is more important than FN.

$$\beta = 0.5$$

$$F - 0.5\ Score = (1 + 0.25)\frac{P * R}{(0.25) * P + R}$$

(3) If FN >> FP

$$F_2\ Score = (1+4)\frac{P * R}{(4 * P + R)}$$