M T W T F S S

- **Linear models for classification**

Linear models are also extensively used for classification.

Two most common linear classification algorithms are:

1) Logistic Regression
2) Linear SVM.

Both algorithms use regularization to overcome overfitting.

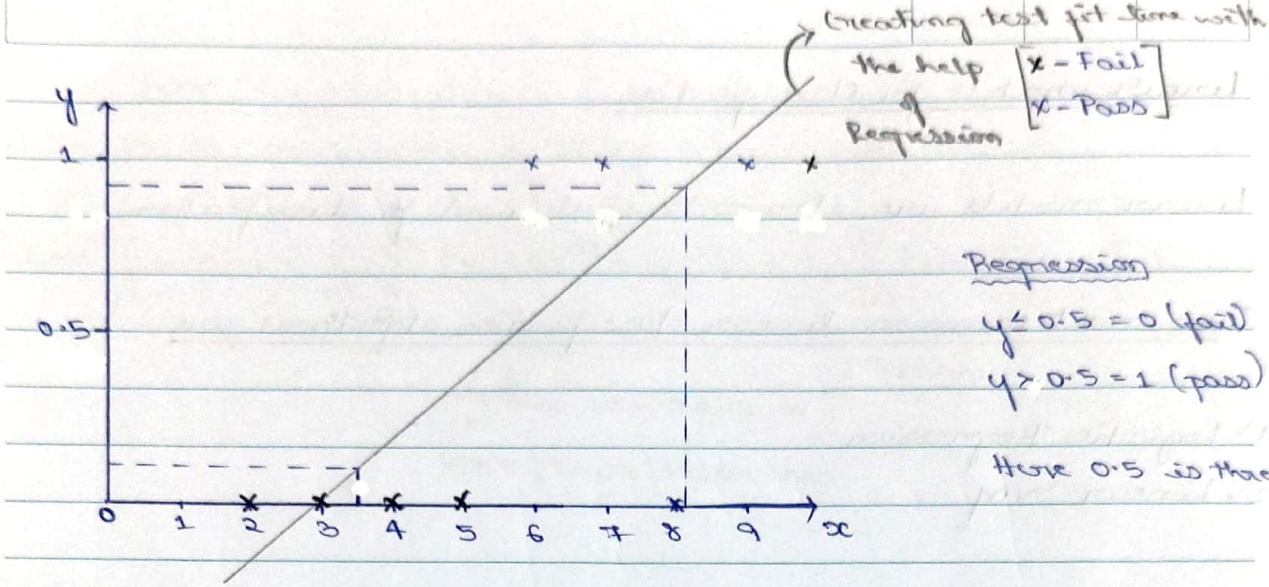
Logistic Regression.

Despite its name, Logistic Regression is a classification problem (algorithm) and not a regression algorithm, and it should not be confused with Linear Regression.

- Consider, the dataset with features 'study hours' and 'output (Pass=1/Fail=0)'

Study hours	o/p (Pass/Fail)
2	Fail
3	Fail
4	Fail
5	Fail
6	Pass
9	Pass
8	Fail ⇒ Exception/outlier.
7	Pass

1) Can we solve this problem using Regression?



Regression

$y \leq 0.5 = 0$ (fail)

$y > 0.5 = 1$ (pass)

Here 0.5 is threshold

New data points

If $x = 3.5$ then $y = ?$ (fail)

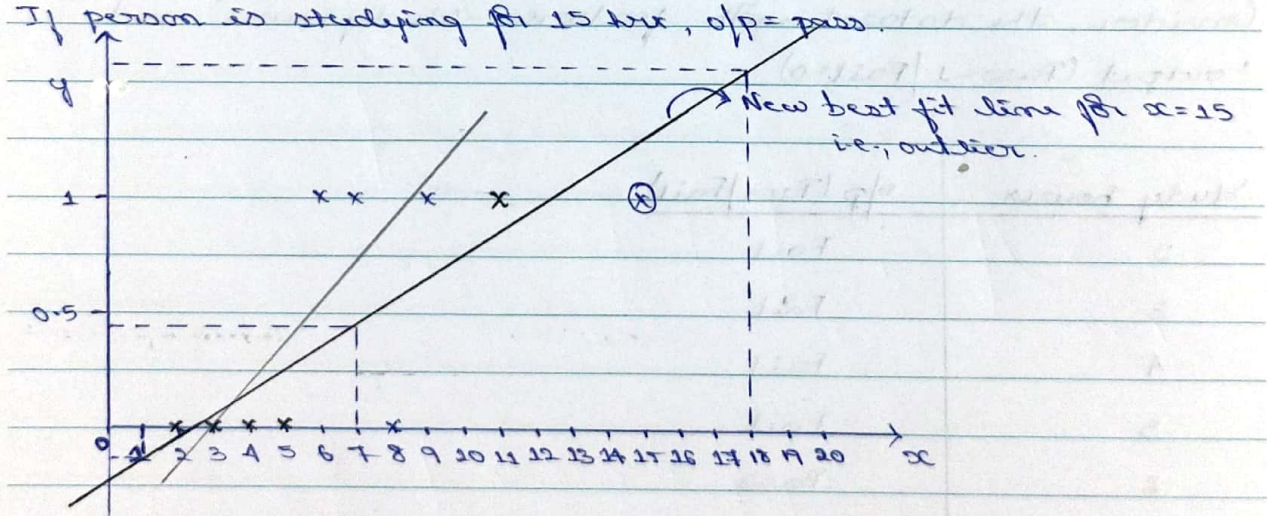
If $x = 8.1$ then $y = ?$ (Pass)

So, we are able to solve this problem using linear regression

Then why Logistic Regression?

Now, consider adding more data points to the model.

If person is studying for 15 hrs, o/p = pass.



⇒ Now, find the o/p

If $x = 7$, then $y = ?$ (fail)

But, initially it gave pass for $x = 7$, due to outlier there is a change in the test fit line.

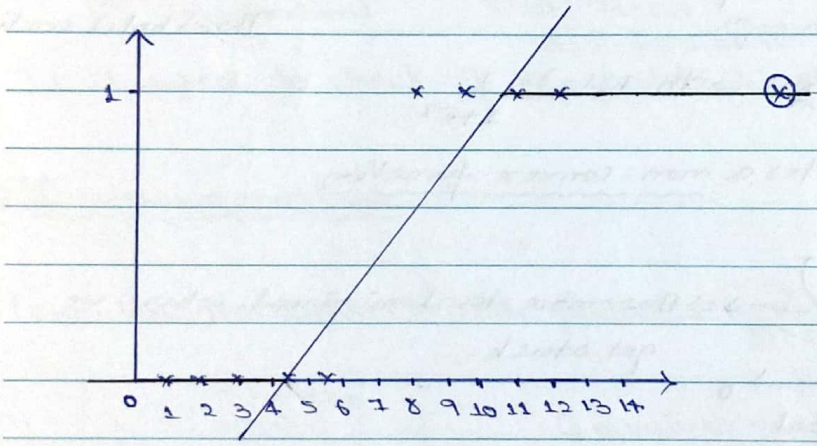
i.e., The model is predicting wrong.

M T W T F S S

If a person is studying for 18 hrs, o/p = greater than 1.
1 hr, o/p = less than 0.

Now, $y > 1$ & $y < 0$ is not been handled
Even though there are outliers, it should be able to give the correct o/p indicating either pass(1) or fail(0) within that given range. This is possible only in Logistic Regression

- We have to cut the best fit line such that everything is b/w 0 to 1 so that it does not get affected by the outliers.



Sigmoid Activation Func.
This func. is responsible for squashing/cut the best fit line.

1) Calculating best fit line using $h_\theta(x) = \theta_0 + \theta_1 x = z$
2) Then we apply 'Sigmoid Activation Func.' on that for squashing it, so that o/p ranges b/w 0 & 1.

Sigmoid func. $= \frac{1}{1+e^{-z}}$; where $z = \theta_0 + \theta_1 x$

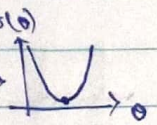
↳ give any value for z, it always gives value b/w 0 to 1 here

WKT, Linear Regression for Cost function:

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

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

↓ gives
Convex function →



↓
1 global minima.

Logistic Regression cost function

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

1) Best fit line

2) Sigmoid function on (Best fit line)

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

Let $z = \theta_0 + \theta_1 x$.

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

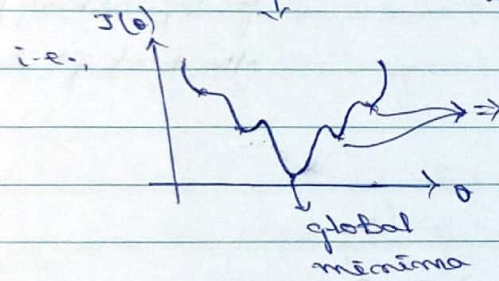
$h_\theta(x) = \sigma(z)$

$$h_\theta(x) = \frac{1}{1+e^{-z}}$$ → gives o/p b/w 0 to 1 (1)

i.e., $< 0.5 \Rightarrow 0$, $> 0.5 \Rightarrow 1$ — Threshold condition

There is a problem with this eqn.

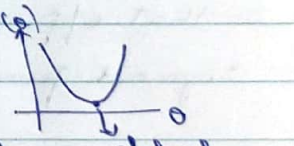
i.e., The cost function with $h_\theta(x) = \frac{1}{1+e^{-z}}$ creates a non-convex function



i.e., ⇒ has more local minimal, where we get stuck.

This problem is fixed by changing the cost function

i.e., We use Log Loss cost function → creates convex function



※ interview concept

- **Log Loss cost function**

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

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

y = Truth value, 0 (or) 1

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

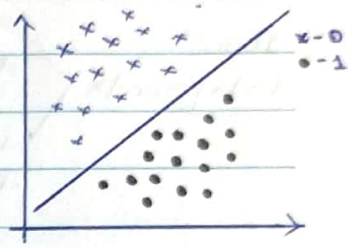
Remaining steps are the same

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

Convergence Algorithm

i.e, Repeat until convergence $\Rightarrow$ { $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ } $j=0$ and 1.

- Performance Metrics (for classification)
  - [a] Confusion matrix
  - [b] Accuracy
  - [c] Precision
  - [d] Recall
  - [e] F-beta score



It is used to check if the model is performing well or not

[a] Confusion matrix

Consider, the dataset with features f_1, f_2 and o/p

f_1	f_2	True value o/p $\Rightarrow y$	$\hat{y}$	
-	-	0	1	×
-	-	1	1	✓
-	-	0	0	✓
-	-	1	1	✓
-	-	0	1	×
-	-	0	1	×
-	-	1	0	×

$\hat{y}$ = values predicted by model

Predicted ($\hat{y}$) \ Actual (y)	1	0
1	3	2
0	1	1

- For binary classification - 2×2 matrix

Predicted ($\hat{y}$) \ Actual (y)	1	0
1	TP	FP
0	FN	TN

TP - True Positive
TN - True Negative
} True = correct match

FP - False Positive
FN - False Negative
} False = Wrong match

M T W T F S S

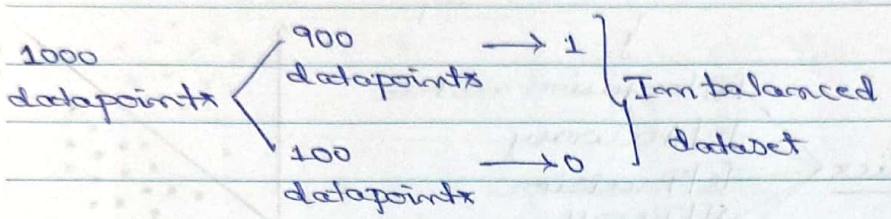
[b] Accuracy

For the previous dataset:

$$Acc = \frac{3+1}{3+2+1+1} = \frac{4}{7} = 0.57 \approx 57\%$$

$$Acc = \frac{TP+TN}{TP+FP+FN+TN}$$

- Consider, the dataset with binary classification having 1000 datapoints



Now, if a dumb model is created in such a way that it gives O/p as 1 then Accuracy

90% ← But it's a bad model ← Only accuracy is not sufficient ← we need to consider Precision, Recall, F-beta score

[c] Precision

Out of all the actual values (TP & FP), how many are correctly predicted (TP)

$$Precision = \frac{TP}{TP+FP}$$

	1	0
1	TP	FP
0		

Actual

Predicted

- In precision we try to focus on FP
- Main aim is to reduce FP

Ex: Problem Statement, Mail is Spam & Ham.

	S	H
S	TP	FP
H		

Actual

Predicted

If, Actual	Predicted	
Spam	Spam	– Good (TP)
Spam	Ham	– we will receive mail it does not have ham. (FN)
Ham	Spam	– we will not receive important mail (FP)

Critical Problem

∴ Focus on reducing FP

[d] Recall

Out of all the predicted values (TP, FN), how many are correctly predicted (TP)

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

Ex: Problem statement, Person has diabetes or not.

Actual	Predicted	
Diabetes	Diabetes	Good (TP)
Diabetes	No Diabetes	Wrong prediction leads to health issue (FN)

Actual \ Predicted	1	0
1	TP	
0	FN	

- In recall we try to focus on FN
- Main aim is to reduce FN.

[e] F-Beta Score

- Consider, Problem statement, if the stock market is going to not crash or crash the next day.

Actual	Predicted	
Crash	Not Crash	(FP) ₹ - Affects companies.
Not Crash	Crash	(FN) ₹ - Affects consumers

Actual \ Predicted	NC	C
NC	TP	FP
C	FN	TN

- Here both FP & FN are important

$$\text{F-Beta} = \frac{(1+\beta^2) * \text{Precision} * \text{Recall}}{(\beta^2 * \text{Precision}) + (\text{Recall})}$$

case 1: If both FP & FN are important; $\beta = 1$.

$$\therefore \text{F-1 score} = 2\,\frac{P*R}{P+R}$$

case 2: If FP is more important than FN; $\beta = 0.5$

$$\therefore \text{F-0.5 score} = \frac{1.25*P*R}{(0.25*P)+R}$$

case 3: If FN is more important than FP; $\beta = 2$

$$\therefore \text{F-2 score} = \frac{5*P*R}{(4*P)+R}$$