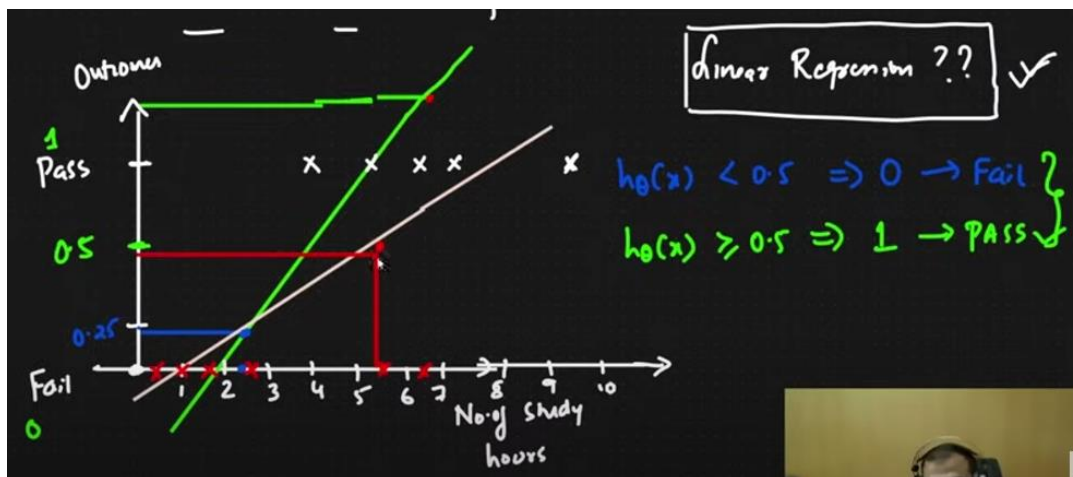
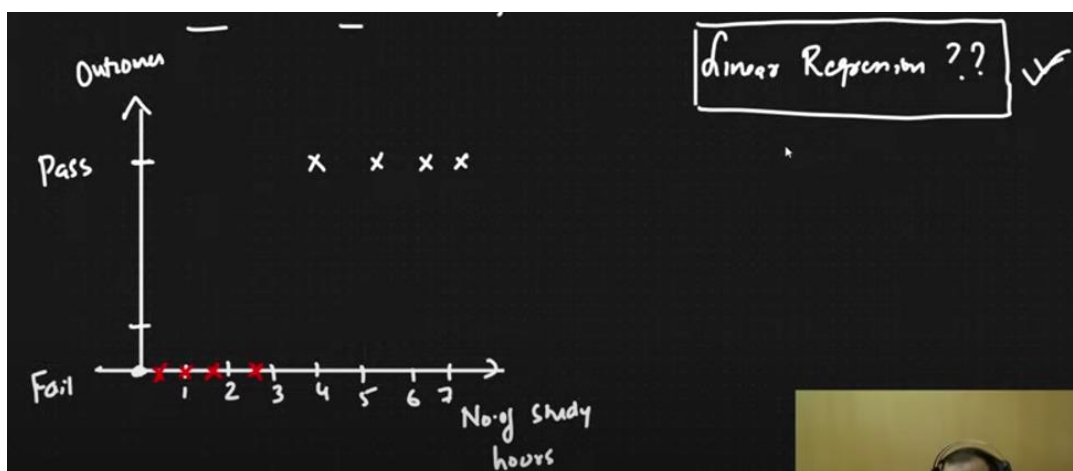


Logistic Regression (classification) → Binary classification

No-of study	No-of play	P/F
—	—	P
—	—	F

1) Unlike, linear regression, here our dependent variable (target variable) is classified in no. of categories! In the above example, it's binary: Pass & Fail!



Over here, we can observe that, linear regression can't be worked on the logistic regression because:

When there's a new output and we draw a new line, the intersection meets at a particular pts which gives us the output FALSE but in reality, it's TRUE!

Also, it can be greater than 1 and lesser than 0 (-ve), which cause us the problems, should be 0 to 1.

Hence, linear regression won't work! Then? Will squash (straight line) called as **Sigmoid function**!

2) Decision boundary in the case of logistic regression:

Decision Boundary Logistic Regression


$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 - \dots + \theta_n x_n$$
$$\boxed{h_{\theta}(x) = \theta^T x}$$

1) create best fit line,

2) squash (straight line on both the side 0 and 1)

3) sigmoid activation function

Math:

$h_{\theta}(x) = \theta_0 + \theta_1 x_1$ 

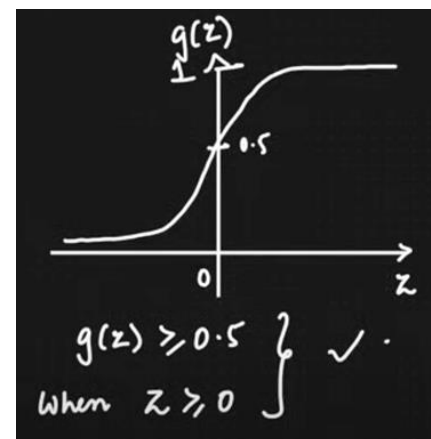
$$h_{\theta}(x) = g(\theta_0 + \theta_1 x_1)$$

Let $z = \theta_0 + \theta_1 x_1$

$$h_{\theta}(x) = g(z)$$

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

$$\boxed{h_{\theta}(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1)}}}$$



3) Why it's called logistic regression?

Because using regression we form a best line and due to sigmoid function, we can solve the error, making the line squash; combined it's called logistic regression!

Training Set

$$\{(x^1, y^1), (x^2, y^2), (x^3, y^3) \dots (x^n, y^n)\}$$

$$y \in \{0, 1\} \rightarrow \text{2 o/p}$$

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

4) When, $\theta_0 = 0$;

Change parameter θ_1 in such a way that, we get the best fit line and apply sigmoid activation function.

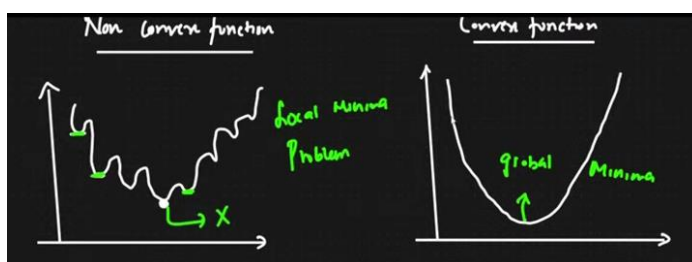
Will find the cost function:

Cost function

Linear Regression $J(\theta_1) = \frac{1}{m} \sum_{i=1}^m \frac{1}{2} (h_{\theta}(x^i) - y^i)^2$

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

replace, $h_{\theta}(x)$ in the above (linear regression) equation **but wait**, we cannot use this equation as $h_{\theta}(x)$ is changing (can take continuous value (any number) but not binary, i.e: $y = 0 / 1$) and so it's non-convex function equation due to which it has lots of local minima. Reaching global minima will be difficult!



Reference!

Then? 5) Will do the following:

$$\text{Cost} = -\frac{1}{m} \sum_{i=1}^m [y \log(y_{\text{pred}}) + (1-y) \log(1-y_{\text{pred}})]$$

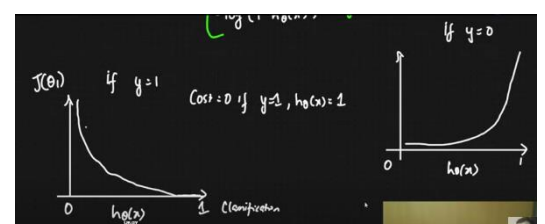
$$y_{\text{pred}} = \sigma(w^T x + b)$$

$y=0$ | error = $-1 \times \log(1-y_{\text{pred}})$

$y=1$ | error = $-\log(y_{\text{pred}})$

a)

$$J(\theta_1) = \begin{cases} -\log(h_{\theta}(x^i)) & y=1 \\ -\log(1-h_{\theta}(x^i)) & y=0 \end{cases}$$



b)

c) final cost function:

$$Cost(h_0(x^i), y) = -y \log(h_0(x^i)) - (1-y) \log(1-h_0(x^i))$$

cost function & loss function with no. of parameters will always be the same!

d) solves the errors w.r.t logistic regression: (log likelihood)

$$J(\theta) = -\frac{1}{2m} \sum_{i=1}^m (y^i \log(h_0(x^i)) + (1-y^i) \log(1-h_0(x^i)))$$

↓
cost

$$h_0(x^i) = \frac{1}{1+e^{-\theta_0 x}}$$

repeat until convergence

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} (J(\theta))$$

6) Performance Matrix: (Classification problems)

x_1	x_2	Actual y	Output \hat{y}
-	-	0	1
-	-	1	1
-	-	0	0
-	-	1	1
-	-	1	1
-	-	0	1
-	-	1	0

To find the accuracy and other terminology of this particular model will perform: Confusion matrix!

Pred \ Actual	1	0
1	3	2
0	1	1

Predicted \ Actual	1	0
1	TP	FP
0	FN	TN

a) Accuracy: $\frac{TP+TN}{Total} = \frac{4}{7} = 0.57 = 57\%$

Try to reduce FP & FN!

$0 \rightarrow 900$ $1 \rightarrow 100$	Imbalanced Dataset	$\rightarrow \left\{ \text{Model} \rightarrow 0 = \frac{900}{1000} = 90\% \right\} \rightarrow \text{Biased Data}$
$0 \rightarrow 600$ $1 \rightarrow 400$	Balanced Data	

Need to perform other terminologies too!

b) Precision: $\frac{TP}{TP+FP}$ (blue)

Example: Spam Classification!

GOAL: Identify which mails aren't spam & are important. Focus on FP; actual spam, predicted imp.!

		1	0	Actual
1	TP	FP		
0	FN	TN		

c) Recall: $\frac{TP}{TP+FN}$ (red)

Example: Has cancer or not!

A person has detected that he/she has cancer but when we test them again, we found they don't have the cancer! Luckily, due to the repeat test, we stated they don't have cancer, just think what if on the basis of actual result, they have been started with the cancer treatment? Hell no!! They might instead get the cancer if they don't have it! So, focus more on FN, actual true, predicted no!

		1	0	Actual
1	TP	FP		
0	FN	TN		

d) F1 -Score / F1 - Beta:

Example: Tomorrow stock market is going to crash

Here, both are important precision and also recall!

$$\begin{aligned} \underline{F\text{-Beta}} &= (1 + \beta^2) \frac{\text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \\ \beta = 1 &= (1 + 1) \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= \frac{2 (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \end{aligned}$$

F1 – Score is like:

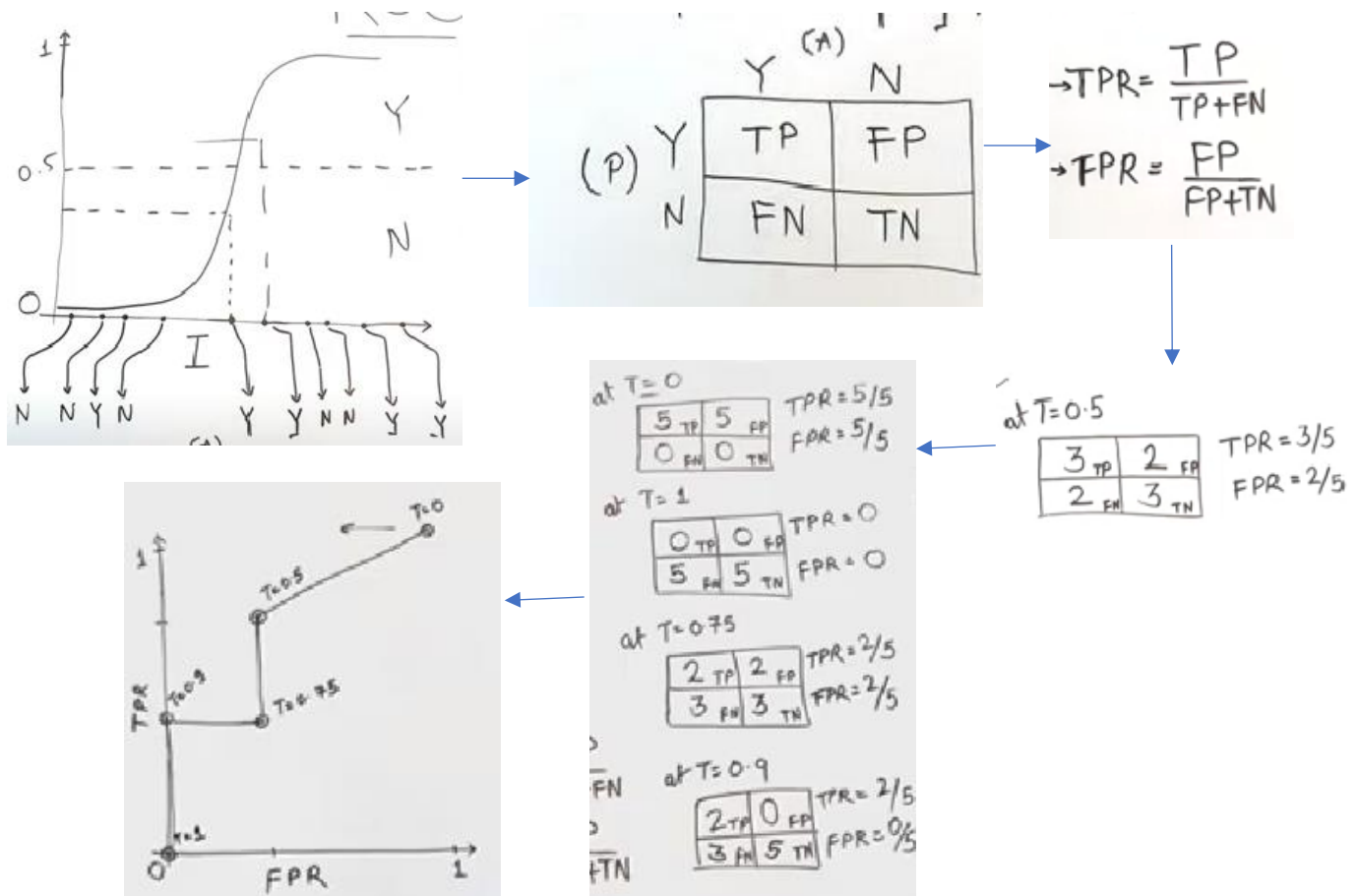
$$\begin{aligned} &\underline{\text{Harmonic Mean}} \\ &= \frac{2xy}{x+y} \end{aligned}$$

when FP is important, will decrease β to 0.5, can say it's F0.5 score;

when FN is important, will increase β to 2, can say it's F2 score!

ON THE GIVEN PROBLEM STATEMENT, WILL GET TO KNOW WHICH TERMINOLOGY IS IMPORTANT

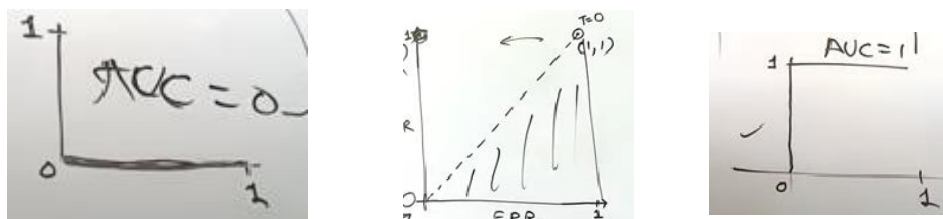
e) ROC (Receiver Operating Characteristic):



ROC - CURVE

To get the best prediction, TPR should be more and FPR should be less!

f) AUC (Area Under Curve): To get the perfect classifier, AUC should be = 1; worst AUC = 0!



Worst

50-50

Best

In simple words, ROC & AUC helps us to understand wheatear our model is the best or not.

Reference:

[1\) Logistic Regression from the scratch \(47:30 mins onwards\)](#)

[2\) Logistic Regression Cost Function](#)

[3\) ROC Curve Part 1](#)

[4\) ROC Curve Part 2](#)

[5\) AUC Curve](#)