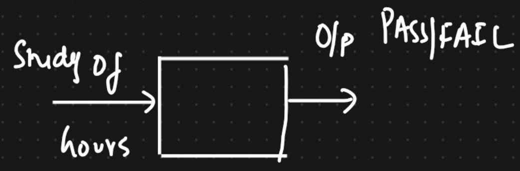
Logistic Regression (Binary classification)

Dataset

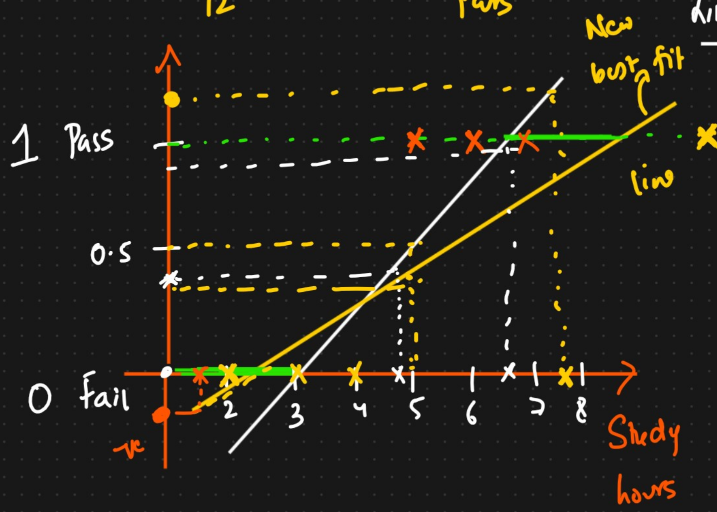
Study hours	Pass/Fail O/p {Binary categories}
2	Fail
3	Fail
4	Fail
5	Pass
6	Pass
7	Pass
12	Pass



Logistic Regression

0 to 1

Linear Regression



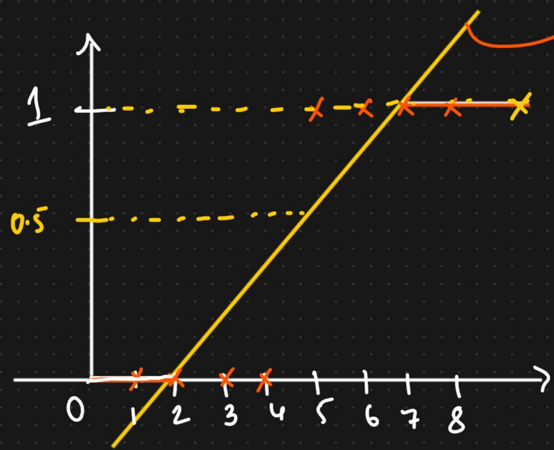
$$\left\{ \begin{array}{l} \le 0.5 \Rightarrow 0 \\ > 0.5 \Rightarrow 1 \end{array} \right\}$$

$$\left\{ \begin{array}{l} > 1 \text{ and} \\ < 0 \end{array} \right\}$$

Why we cannot use Linear Regression for Classification?

1. Outlier {Best fit line Change}
2. >1 and <0 {Squash line}

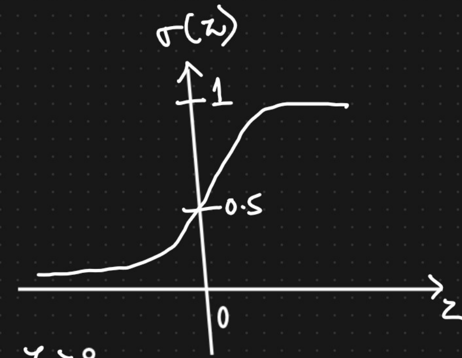
How Logistic Regression Solves Classification Problem



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

⇓

Best fit line



$z \geq 0$

$\sigma(z) \geq 0.5$

Sigmoid Activation

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

$z = \theta_0 + \theta_1 x_1$

⇑

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

$$= \sigma(z)$$

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

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

⇒ Logistic Regression hypothesis

$$z = \theta_0 + \theta_1 x_1$$

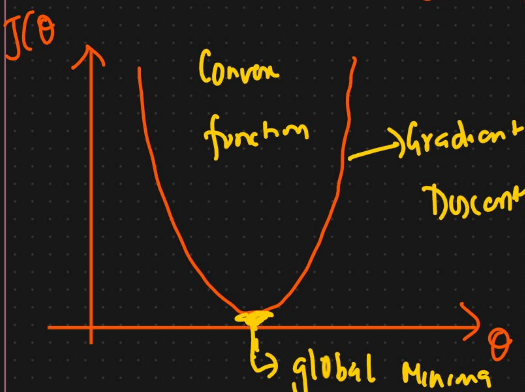
Linear Regression Cost function

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

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

⇓

Convex function



Logistic Regression Cost function

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

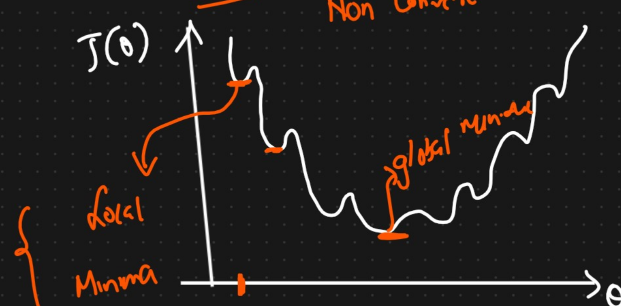
Non Convex function

⇓

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

Sigmoid

Non Convex



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

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

$\Downarrow$

Lets Denote $Cost(h_\theta(x)^{(i)}, y^{(i)})$

$\Downarrow$

{ Log Loss }

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

Convex function

$\Downarrow$

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

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

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

Convergence Algorithm

Repeat

{ $j = 0$ and 1

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

}

Performance Metrics, Accuracy, Precision, Recall And F-Beta

Logistic Regression

R squared

Adjusted R squared

Topics to be covered

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

Dataset

O/P — pred by model

x_1	x_2	y	$\hat{y}$
–	–	0	1
–	–	1	1
–	–	0	0
–	–	1	1
–	–	1	1
–	–	0	1
–	–	1	0
–	–		

① Confusion Matrix

Actual Values

	1	0
1	3	2
0	1	1

Predicted Values

	1	0
1	TP	FP
0	FN	TN

Actual

Predicted

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

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

$$= \frac{4}{7}$$

④ Dataset — Binary classification

↳ 1000 datapoints → 900 → 1, 100 → 0 } Imbalanced Dataset

⇓

90% Accuracy

④ **Precision** $= \dfrac{TP}{TP + FP\downarrow}$ } Out of all the actual value how many are correctly predicted

Predicted \ Actual	1	0
1	TP	FP
0	FN	TN

② **Recall** $= \dfrac{TP}{TP + FN\downarrow}$ ⇒ Out of all the predicted value how many are correctly predicted

Usecase ①

Spam classification

Mail → Spam
Model → Spam } Good

	Spam 1	Not Spam 0 (Actual)
Spam 1	TP	FP ✓ ⇓
Not Spam 0	FN	TN

Mail → Not Spam
Model → Spam } Blunder

⇓

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

Usecase 2

To predict whether person has diabetes or not

✓ Truth → diabetes
✓ Model → Doesn't diabetes } Blunder

	Diabetes	No Diabetes
Diab	TP	FP
No Diab	FN	TN

↓ ⇓ usecase of diseases

Recall $= \dfrac{TP}{TP + \boxed{FN}\Downarrow}$

Truth → diabetes
Model → " } good

Truth → Not diabetes
Model → Diabetes

$\Rightarrow$ 2nd opinion
$\Downarrow$
check

Assignment

(*) Tommorow the stock market will crash or not

Reducing FP↓ or FN↓

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

① If FP & FN are both important
$\beta = 1$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

$\Rightarrow$ Harmonic Mean

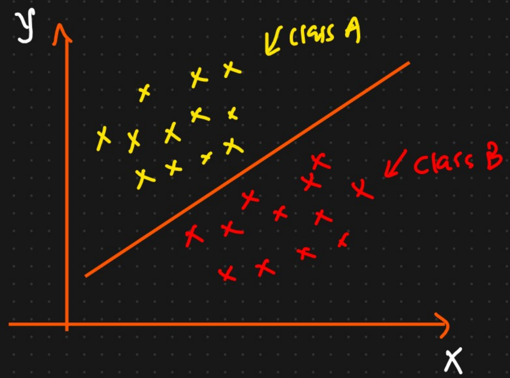
② If FP is more important FN
$\beta = 0.5$

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

③ If FN >> FP
$\beta = 2$

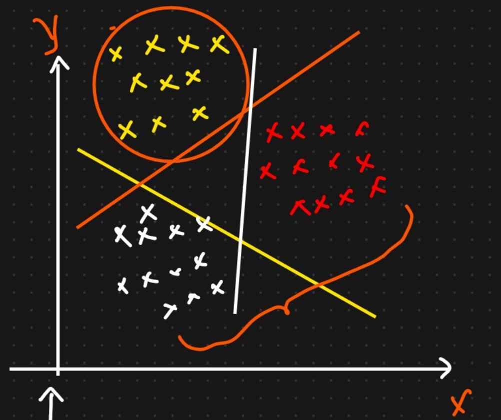
$$\boxed{\text{F 2 Score} = (1+4) \frac{P \times R}{P + R}}$$

Logistic Regression (One Versus Rest)



① Binary Classification

One Versus One



Multiclass Classification {Logistic Regression}

→ M_1 → Binary classification
→ M_2 → Binary classification
→ M_3 → Binary classification

One Versus Rest (OvR) → Logistic Regression

O_1, O_2, O_3

f_1	f_2	f_3	O/p	$\boxed{O_1}$	O_2	$\boxed{O_3}$
–	–	–	O_1	1	0	0
–	–	–	O_2	0	1	0
–	–	–	O_3	0	0	1
–	–	–	O_1	1	0	0
–	–	–	O_3	0	0	1
–	–	–	O_2	0	1	0

Model

M_1 ← I/p $\{f_1, f_2, f_3\}$ O/p $\{O_1\}$
M_2 ← I/p $\{f_1, f_2, f_3\}$ O/p $\{O_2\}$
M_3 ← I/p $\{f_1, f_2, f_3\}$ O/p $\{O_3\}$

0.55 → O/P = $\boxed{O_3}$ → category 3

$[0.25, 0.20, \boxed{0.55}]$

↑ M_1 ↑ M_2 ↑ $\boxed{M_3}$ ↑ O/p

New Test Data
→ M_1 → 0.25 ✓
→ M_2 → 0.20 ✓
→ M_3 → 0.55 ✓