Classification

Email - Spam / Not
Online tran - fradulent / Not
Exam - Pass / Fail
Loan - Accepted / No
Gender - Male / Female
Healthcare - Cure / Not
tumour - Cancer / Not
Sales - Buy / Not

Target Regression - Continuous
What if it is discrete ?

Logistic Regression ✓

Classification Algorithm

Output Linear Regression

$(-\infty \quad , \quad +\infty)$ | $e^{-x} \Rightarrow$

x	-1000	-2	-1	0	1	2
e^{x}	5.0	0.135	0.367	1	2.71	7.3

range $(0, \quad +\infty)$

$5565 \Rightarrow \frac{p}{p+1} = 0 > x < 1$

$(0 \quad , \quad 1)$

$\boxed{y = mx + b}$ linear Regression

$y = (-\infty, \infty) \Rightarrow (0, 1)$

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

$$\boxed{\frac{e^{(mx+b)}}{e^{(mx+b)} + 1} = P}$$

probability

$(0, 1)$

$$\boxed{P = \frac{e^y}{e^y + 1}}$$

$$P = \frac{e^y}{e^y + 1}$$

$y = \boxed{(\quad)}$

$P(e^y + 1) = e^y$

$Pe^y + P = e^y$

$$pe^y - e^y = -p$$
$$e^y - pe^y = p$$
$$e^y(1-p) = p$$
$$e^y = \frac{p}{1-p}$$

e^y
$\downarrow$
Anti log

Apply log on both side

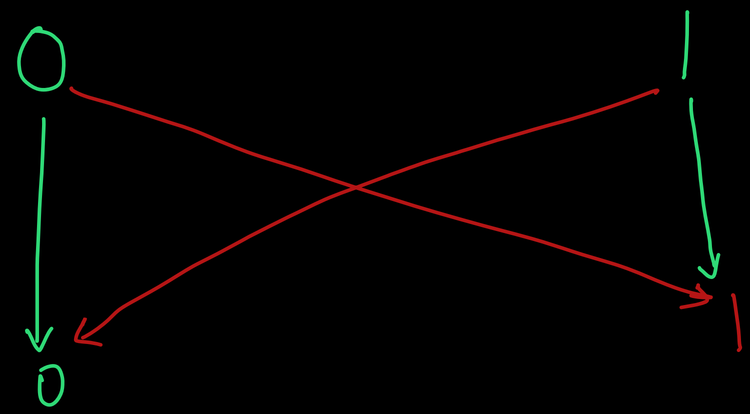
$$\log(e^y) = \log\left(\frac{p}{1-p}\right)$$
$$y = \log\left(\frac{p}{1-p}\right)$$

$$\boxed{\log\left(\frac{p}{1-p}\right) = mx + b}$$

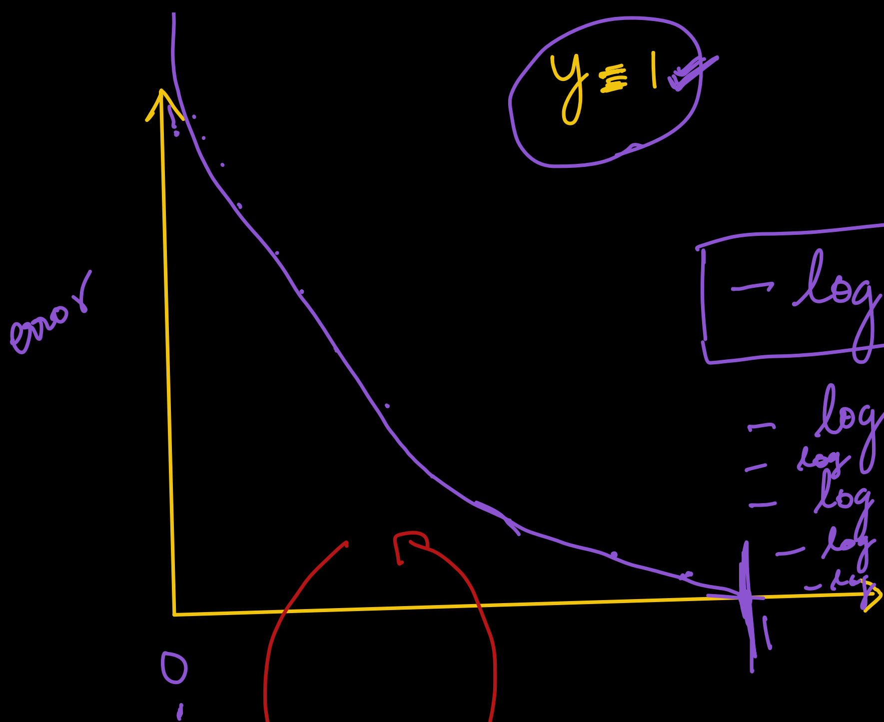
p = probabity of success
1-p = probabity of failure

$$\left.\frac{\text{p. Success}}{\text{p. Failure}}\right| \Rightarrow \text{Odds}$$

$$\log(\text{odds}) = mx + b$$



		Error
$0 \longrightarrow 0$	$\rightarrow$	0
$0 \longrightarrow 1$	$\rightarrow$	∞
$1 \longrightarrow 1$	$\rightarrow$	0
$1 \longrightarrow 0$	$\rightarrow$	∞



$y = 1$

$\boxed{-\log(\hat{y})}$

$-\log(1) = 0$ ✓

$-\log(0.9) = 0.04$

$-\log(0) = \infty$

$-\log(0.7) = 0.154$

$-\log(0.1) = 1$

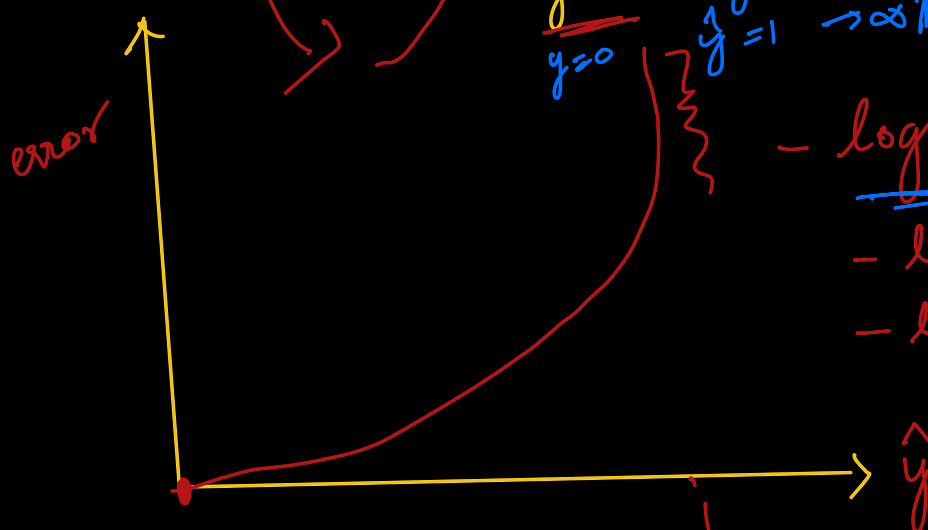
$y = 0 \rightarrow \hat{y} = 0 \rightarrow 0$

$y = 0 \quad \hat{y} = 1 \rightarrow \infty$

$-\log(1 - \hat{y})$

$-\log(1-1) = \infty$

$-\log(1-0) = 0$



Cost function

$y = 1$: $-\log(\hat{y})$ ✓

$y = 0$: $-\log(1 - \hat{y})$ ✓

Cost function

$\boxed{\text{log loss}}$

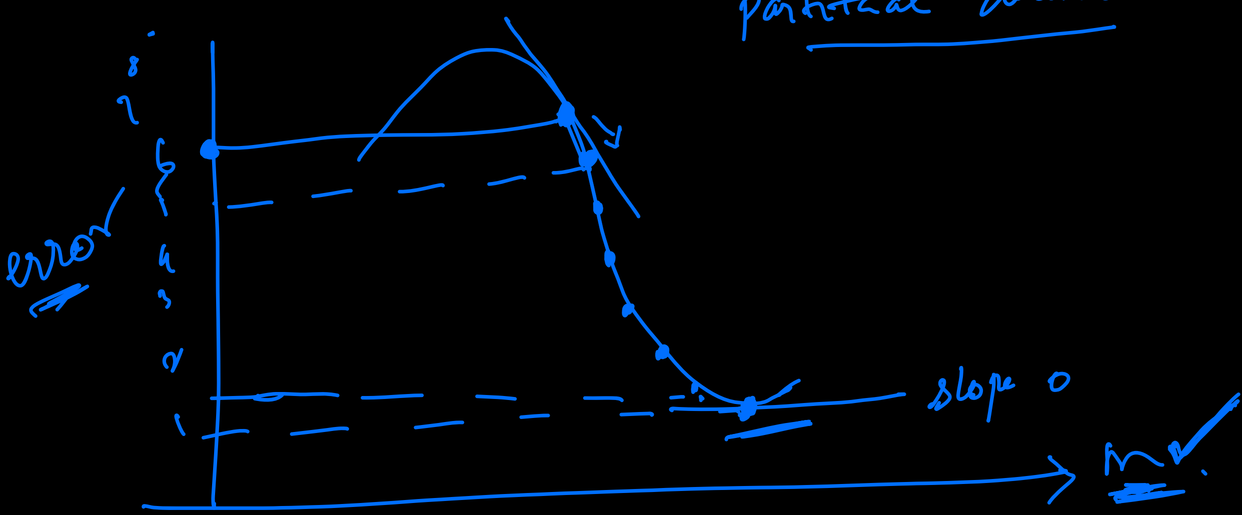
$$\text{Cost} = \left[-y\log(\hat{y}) - (1-y)\log(1-\hat{y}) \right]$$

$y = 1$ $\qquad\qquad$ $y = 0$

$$-y\log(mx+b) - (1-y)\log(1-(mx+b))$$

iterative Algorithm

partital durrative



Confusion Matrix

$0 \to 0$

$0 \to 1$

$1 \to 1$

$1 \to 0$

		Actual Values	
		Yes (Fraud)	No (Not Fraud)
Predicted Values	Yes (Fraud)	True Positive TP	False Positive FP
	NO (Not Fraud)	False Negative FN	True Negative TN

False Positive ⇒ Falsely predicting Negative as positive

y	$\hat{y}$	output for threshold 0.6	0.4
0 ✓	0.5	0	1
1 ✓	0.9	1	1 — ✓
0 ✓	0.7	1	1
1 ✓	0.7	1	1 — ✓
1 ✓	0.3	0	0 ×
0 ✓	0.4	0	1
1 ✓	0.5	0	1 — ✓

3/4

3/6

1/2

	Actual 1 Pos	Actual 0 Neg	
Predicted 1 Pos	2 (TP)	1 (FP)	3
Predicted 0 Neg	2 (FN)	2 (TN)	4
	4	3	7

Metrics

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

$= \dfrac{4}{7}$

100 record

10 8 98 record negative Not cancer

2 record positive Cancer

100 Not Cancer

accuracy = $\frac{98}{100}$ = 98%

Recall :

$$= \frac{TP}{TP + FN}$$ (Actual)

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

Precision:

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

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

F_Score

Harmonic mean of Precision & Recall

$$\frac{2\,P \cdot R}{P + R} \qquad \frac{P+R}{2}$$