Logistic Regression - 2

13 August 2020 07:09

- Logistic regression is used for solving classification problems - two class classification
- Logistic regression is a linear model, it expects the features to have good linear relationship with the label

$\hat{y} = mx + c$

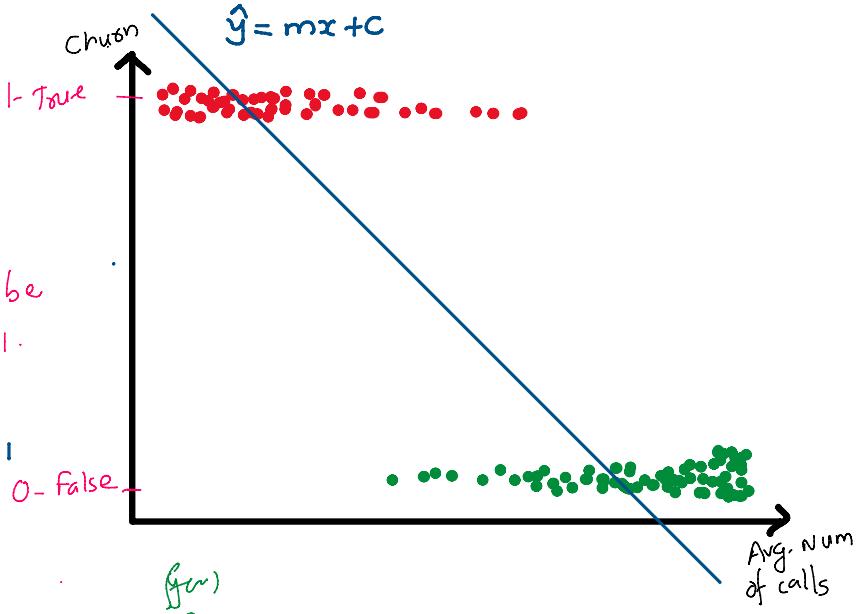
Sigmoid function / logit function

$f(x) = \frac{1}{1+e^{-x}}$

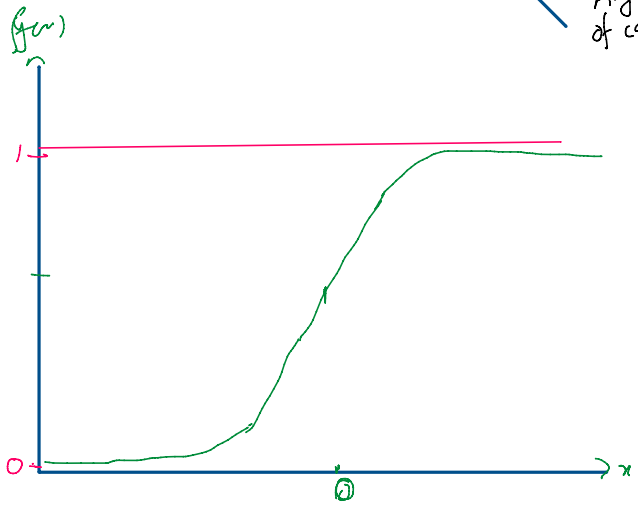
for any value of x, $f(x)$ will always be b/w 0 to 1.

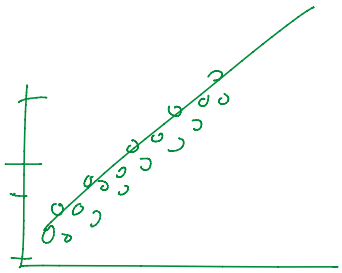
$\frac{1}{1+e^{-\hat{y}}} = P$ If $P \geq 0.5 \Rightarrow$ class 1

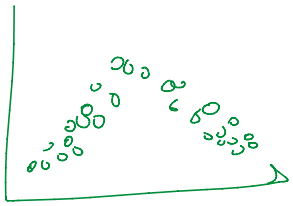
If $P < 0.5 \Rightarrow$ class 0

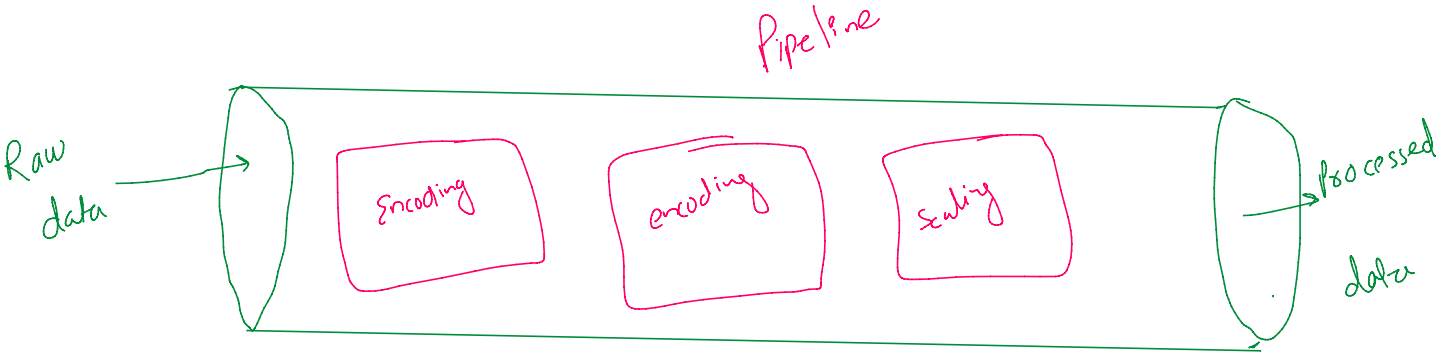


$f(x) = \frac{1}{1+e^{-x}}$









	Actual	Prediction
	1	1
	1	0
	1	0
	0	0
	0	0
	0	1
	0	1
	0	0
	1	0
	1	1
	0	1
	1	0
	1	0
	0	0
	1	0

$7/15 = 47.5\%$

$R = 2/8$ $= 25\%$

100 Observation
→ 90 – 0 – Not leaving
10 – 1 – leaving

Damaged 10 sites → 5 sites
3 CR, 2 WCR

R = 3/10

Precision = 10/10

P = 2/5

Prediction model

		Correct
95	– 0 –	90
05	– 1 –	05

Accuracy = 95/100 = 95%

Recall (accuracy of class 1) = $\frac{5}{10}$ = 50%

model -2

~~80 — 0 —~~ Correct

		Correct
82	– 0 –	82
18	– 1 –	10

A = $\frac{92}{100}$ = 92%

R = $\frac{10}{10}$ = 100%

Accuracy – Correctness of model for both class 1 and class 0

Recall – Correctness of model for only class 1.

Actual
100 → 90 – 0 – not leaving
10 – 1 – leaving
(10: Actual num of obs in class 1)

Prediction

		Correct	Incorrect
95	0 – not leaving	88	07
05	1 – leaving	03	02
		91	

(05: obs. predicted as class 1; 03: correct predictions in class 1)

Accuracy = $\frac{88 + 03}{100}$ = 91%

Recall = $\frac{\text{Correct Predictions in class 1}}{\text{Actual num of obs. in class 1}}$ = $\frac{03}{10}$ = 30%

Precision = purity in predictions
= $\frac{\text{correct predictions in class 1}}{\text{total predictions in class 1}}$ = $\frac{03}{05}$ = 60%

0 – 0 – 0
→ 100 – 1 – 10

R = 10/10 = 100%.

Telecom churn
reward

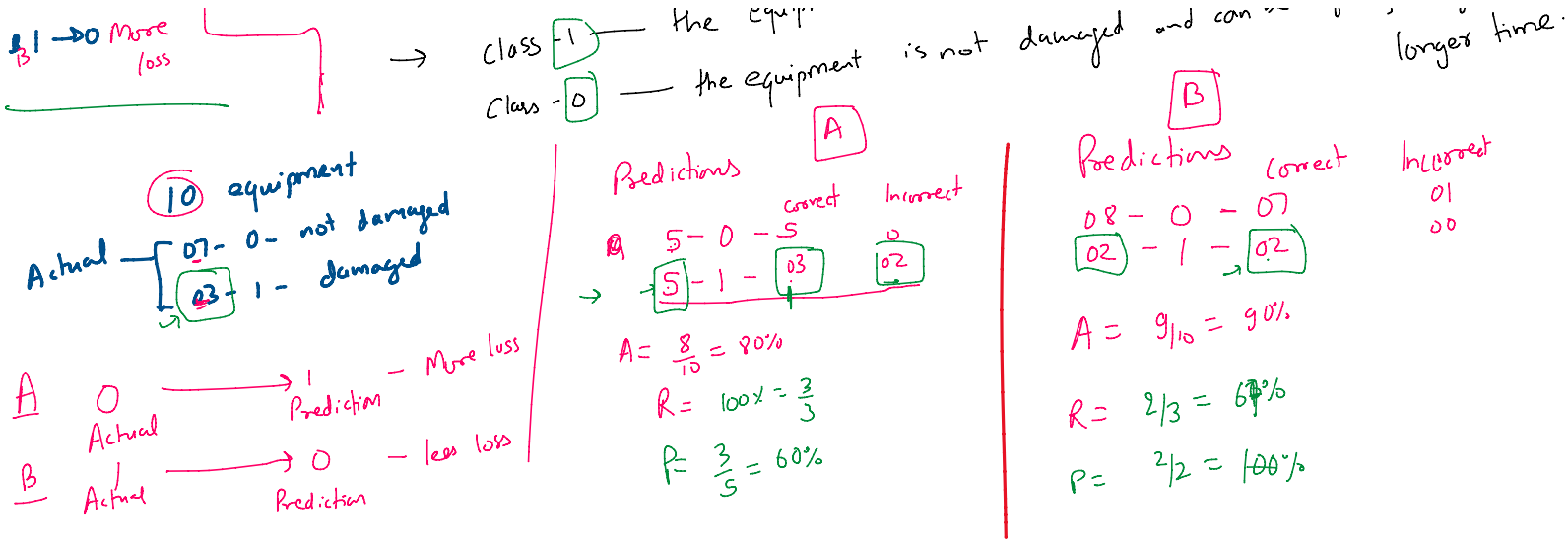
A 0 → 1 less loss
B 1 → 0 More loss

Site equipment — Scrapping

→ Class 1 — the equipment is almost damaged and should be scrapped
~~longer time~~

Class 0 — the equipment is not damaged and can be useful for longer time.

B



Confusion Matrix

Predictions

Actuals →

Confusion Matrix	0 Negative	1 Positive
0	True Negative TN	False Positive FP
1	False Negative FN	True Positive TP

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

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

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

$$\text{F1 score} = \frac{1}{\frac{1}{R} + \frac{1}{P}} = \frac{2 \times P \times R}{P + R}$$