Logistic Regression with Regularization

→ cost function

$$J(\theta_0,\theta_1) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x)) + L_2$$

we Add L_2 Regularization
↳ Reduce over fitting.

we can Add L_1 Regularization
↳ feature selection

we can Add Both $L_1 + L_2$ Regularization.

for L_2 we Add $\lambda \sum_{i=1}^{n} (\text{slope})^2$
for L_1 we Add $\lambda \sum_{i=1}^{n} |\text{slope}|$
} λ = Hyperparameter

C & lambda (λ) Relationship

$$\boxed{C = \frac{1}{\lambda}}$$

* Performance metrics, Accuracy, Precision, Recall and F-Beta.

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

1) confusion metrics

Actual value (y)

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

Actual value (y)

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

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

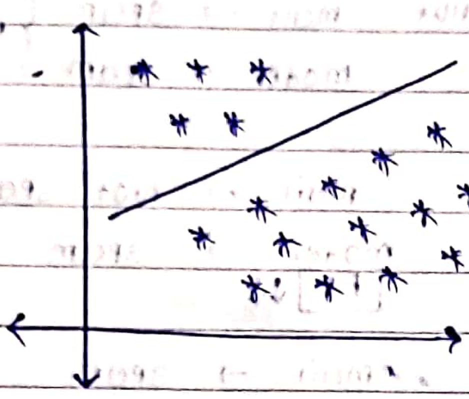
$$= \frac{3+1}{3+2+1+1} = \frac{4}{7} = \boxed{0.57}$$

model Accuracy. $\boxed{57\%}$

1) confusion metrics

2) Accuracy of model.

Page No:
Date:



* dataset → Imbalance dataset

1000 datapoint → $\boxed{900 \to 1}$, → 100 → 0

Dumb model → O/P → 1

Accuracy → 90%

cuz of 900 datapoints

→ In the case of imbalance dataset we can't use Accuracy performance.

→ Here we use Precision or Recall.

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

	1	0
1	TP	TN
0	FN	FP

Here, TP → out of all the actual values how many are correctly predicted.

FP is important
Here FP ↓↓ [Reduce false positive]

Recall $= \frac{TP}{TP + FN}$ → out of All the predicted values How many are correctly predicted with Actual values.

→ Usecase → 1
Spam classification mail → spam / → Not-spam

Actual

	1	0
1	TP	FP
0	FN	TN

Predicted value

mail → spam, model → spam: good

mail → Not spam, model → spam: Blunder [FP] ↓↓

mail → spam, model → Not a spam: Fine [FN]

For these scenario we use precision metrics.

→ use case - 2

To predict whether person Has diabetes or not

Actual → yes, model predict → yes: good

Actual → yes, model predict → NO

Actual → NO, model predict → yes: Blunder predicted value

	1	0
1	TP	FP
0	FN	TN

Actual value

we can Reduce FN so Recall is use.

→ Health desease use RECall is use.

Assignment 1) Tomorrow stock market will crash or Not.

Actual → crash
Predict → crash } good

Actual → crash
Predict → Not crash } Blunder

Actual → Not crash
Predict → crash } $\boxed{FN}$ ↓↓

	1	0 (Actual)
1	TP	FP
0	FN	TN

Predicted

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

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

F1 score $= 2 * \dfrac{\text{Precision} * \text{Recall}}{\text{precision} + \text{Recall}}$ } Harmonic score

If FP is more Imp Than FN !!

$\beta = 0.5$

F 0.5 score $= (1 + 0.25)\; P * R / P+R$

If FN >> FP $\beta = 2$

F_2 score $= (1+4)\;\dfrac{P * R}{P+R}$