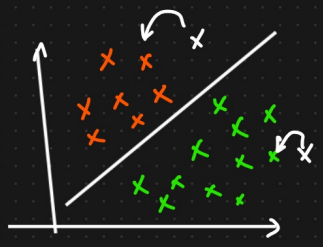
Performance Metrics, Accuracy, Precision, Recall And F-Beta

Topics to be covered

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



Logistic Regression

- In case of Regression
  - R squared
  - Adjusted R squared

Dataset

x_1	x_2	y (o/p)	ŷ (pred by model)
-	-	0	1
-	-	1	1
-	-	0	0
-	-	1	1
-	-	1	1
-	-	0	1
-	-	1	0
-	-		

① Confusion Matrix

	Actual 1	Actual 0
Predicted 1	3	2
Predicted 0	1	1

Predicted values ↓ , Actual values →

	Actual 1	Actual 0
Predicted 1	TP	FP
Predicted 0	FN	TN

A confusion matrix(2x2) is a performance measurement tool used for classification models

TP ;- True Positive	FP :- False Positive
FN:- False Negative	TN :- True Negative

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

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

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

E.g.

① Dataset — Binary classification

↳ 1000 datapoints

- 900 → 1
- 100 → 0

} Imbalanced Dataset

⇓

90% accuracy

But in the real it is not true. In the imbalanced data set, it can give you a completely biased result.

④ Precision = $\frac{TP}{TP+FP}$ ↓

Out of all the actual value how many are correctly predicted

	1	0
1	TP	FP
0	FN	TN

Actual

Predicted

- Focuses on reducing False Positives (FP)
- Useful when False Positives (FP) are costly (e.g., fraud detection).

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

- Focuses on reducing False Negative (FN)
- Important when False Negatives (FN) are costly (e.g., medical diagnosis).

Usecase ①

Spam classification

Mail → Spam
Model → Spam
} Good

	Spam 1	Not Spam 0
Spam 1	TP	FP ✓ ↓↓
Not Spam 0	FN	TN

Actual

Mail → Not Spam
Model → Spam
} Blunder

↓↓

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

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

Recall = $\frac{TP}{TP+FN}$ ↓↓

Usecase of diseases

Truth → diabetes
Model → "
} good

Truth → Not diabetes
Model → Diabetes
} ⇒ 2nd opinion ⇓ check Not blunder

Assignment

① Tommorow the stock market will crash or not

Reducing FP↓ or FN↓

Ans :- Reduce FN, bcoz if predicting Not-crash, actually crash it will be blunder.

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

① If FP & FN are both important

$\beta = 1$

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

$\beta = 2$

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