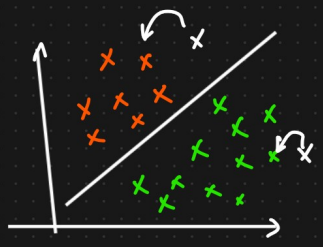
Performance Metrics, Accuracy, Precision, Recall And F-Beta

Topics to be covered

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



Logistic Regression

R squared

Adjusted R squared

Dataset

x_1	x_2	o/p y	pred by model $\hat{y}$
-	-	0	1
-	-	1	1
-	-	0	0
-	-	1	1
-	-	1	1
-	-	0	1
-	-	1	0
-	-		

1. Confusion Matrix

Actual values \ Predicted values	1	0
1	3	2
0	1	1

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

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

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

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

2. Dataset

Binary classification

- 1000 datapoints
  - 900 → 1
  - 100 → 0

Imbalanced Dataset

⇓

90% Accuracy

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

Predicted \ Actual	1	0
1	TP	FP
0	FN	TN

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

Usecase ①

Spam classification

Mail → Spam
Model → Spam
} Good

Spam \ Actual	1	0
1	TP	FP ✓ ↓↓
0	FN	TN

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

Assignment

(*) Tommorow the stock market will crash or not

Reducing FP↓ or FN↓

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

① If FP & FN are both important

$\beta = 1$

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

⇒ Harmonic Mean

② If FP is more important FN

$\beta = 0.5$

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

③ If FN >> FP

$\beta = 2$

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