Metrics for evaluating ML Models

Classification Problem

- Binary
- Probability
  - Cut off is very important to determine hyperparameter

① Confusion Matrix

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

FP → Type 1 Error → $FPR = \frac{FP}{FP+TN}$

FN → Type 2 error = $\frac{FN}{FN+TP}$ FNR

② Accuracy : Accuracy is a ~~mat~~ metrics which is used when your dataset is balanced. If we have a churn and non churn data and churn % is just 10%. and non churn is 90%.. Accuracy is not a nice metrics when your data is baised. Similarly in multiple classification when data is baised one should not use accuracy..

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

(Sensitivity) Recall (TPR) = $\frac{TP}{TP+FN}$ → health

Precision (+ve Pred val) = $\frac{TP}{TP+FP}$ → spam

F_1 score | F Beta (imbalanced dataset) = $(1+\beta^2)\frac{\text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}$

$$f_1 = \left[\frac{2\,\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right]$$ ~~Geometric~~ Harmonic mean.

$\beta = ?$

$\beta = 1$; False positive and false negative are equally important

Precision ↑↑, $0 < \beta < 1$; false positive have more impact than false negative

Recall ↑↑, $\beta > 1$; false negative has more impact than false positive

ROC and AUC Curve (mostly used for binary classification problem)

threshold values [0, 0.2, 0.4, 0.6, 0.8, 1]

y	$\hat{y}_{pred}$	$\hat{y}(0)$	$\hat{y}(0.2)$
1	0.8	1	1
0	0.96	1	1
1	0.4	1	1
1	0.3	1	1
0	0.2	1	0
1	0.7	1	1

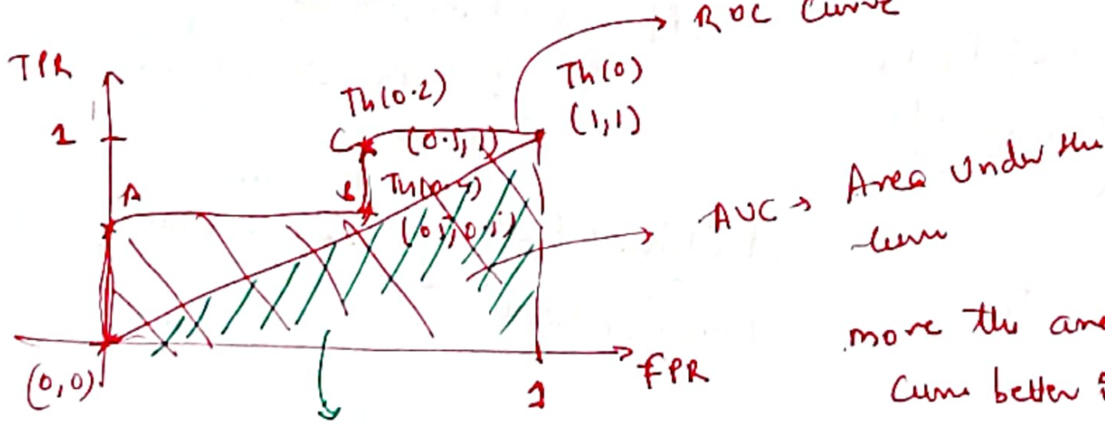
$$TPR = \frac{TP}{TP+FN} = \frac{4}{4+0} = 1$$

$$FPR = \frac{FP}{TN+FP} \Rightarrow \cancel{\frac{4}{4+2} = \frac{4}{6} = 2/3 = 0.66}$$

$$= \frac{2}{0+2} = 1$$

$$TPR = \frac{4}{4+0} = 1$$

$$FPR = \frac{FP}{FP+TN} = \frac{1}{1+1} = 0.5$$



more the area under the curve better is the model

Total Area should always be greater than this green area

$\frac{1}{2} \times 1 \times 1 = 0.5$

Q. What threshold value we should select from this?

- Now go to a domain expert then the expert will let you know what he cares about

True Positive >> false Positive (A)

false Positive >> True Positive (B)

True Positive ≃ false Positive (C)

Bias - Variance of the model (Overfitting | Underfitting)

prefered model = low bias, low variance

very little parameters :- high bias | low variance
very high parameter :- low bias | high variance

high bias → oversimplified → underfit
low bias → over complicated → overfit

high variance → prediction ~~clustered~~ spread across the real target

low varn → prediction clustered near the real target

Irreducible error.

$$Err(x) = \underbrace{\left(E[\hat{f}(x)] - f(x)\right)^2}_{Bias^2} + \underbrace{E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right]}_{var} + \underbrace{\sigma_e^2}_{Irr\ error}$$

Bias, Variance tradeoff

Cross Validation | K-fold Cross validation

In cross validation we run our modelling process of different subsets of data to get multiple measures of model quality.

Ex 1   val   Train

Ex 2   T   val   T   T   T

Ex 3   T   T   val   T   T

Ex 4   T   T   T   val   T

Ex 5   T   T   T   T   val