Classification Metrics 2

26 December 2024 18:16

☆ Accuracy, Precision, Recall, F-1 score etc.

How many such metrics are there?

Sources: [12][13][14][15][16][17][18][19] view · talk · edit

		Predicted condition			
	Total population $= P + N$	Predicted positive (PP)	Predicted negative (PN)	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Actual condition	Positive (P) [a]	True positive (TP), hit[b]	False negative (FN), miss, underestimation	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate type II error [c] $= \frac{FN}{P} = 1 - TPR$
	Negative (N)[d]	False positive (FP), false alarm, overestimation	True negative (TN), correct rejection[e]	False positive rate (FPR), probability of false alarm, fall-out type I error [f] $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$
	Prevalence $= \frac{P}{P+N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$
	Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{PN} = 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
	Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	F_1 score $= \frac{2 PPV \times TPR}{PPV + TPR} = \frac{2 TP}{2 TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times FDR}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$

☆ No need to remember because all of them are generated from TN, FP, FN, TP

☆ Most imp. are: accuracy, precision, recall, F1 score

☆ Other imp metrics are: TPR, FPR, TNR, FNR

Predicted

Actual		N	P
N_a	N	TN	FP
P_a	P	FN	TP
		N_p	P_p

TPR = True Positive Rate

FPR = False " "

TNR = True Negative "

FNR = False " "

☆ **Rule: Always divide by Actual**

$$\boxed{TPR = \frac{TP}{TP + FN}}$$

Recall

$$\boxed{FPR = \frac{FP}{FP + TN}}$$

$TP + FN$ $\qquad\qquad$ $FP + TN$

Recall

$$TNR = \frac{TN}{TN + FP} \qquad\qquad FNR = \frac{FN}{FN + TP}$$

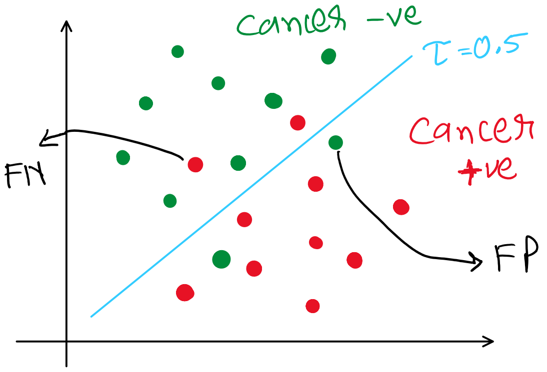
☆ Recap Process of Logistic Regression

$$x_i \rightarrow \underset{\substack{z_i \\ (-\infty,\ \infty)}}{w^T x + w_0} \rightarrow \underset{[0,\ 1]}{\sigma(z_i)} \begin{cases} < 0.5 & \text{class-0} \\ \geq 0.5 & \text{class-1} \end{cases}$$

Threshold $\Rightarrow \tau = 0.5$

Should we always take $\tau = 0.5$? If not, why & when should we take a different τ?

Example - Cancer detection

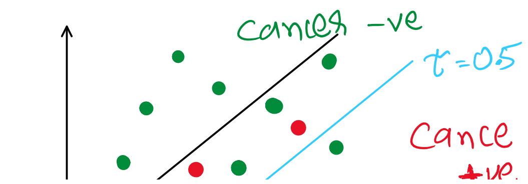


Let's consider FN:

our model has declared these patients as 'cancer -ve' so, they were not recommended any further test / treatment. This may lead their condition worsen. This is bad.

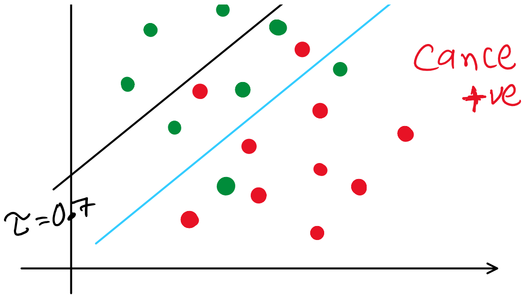
Lets talk about FP now: Our model has declared these patients as 'cancer +ve'. Hence we will recommend them the next round of test. This is not as bad.

In such cases when we want our model to be more sensitive to either FN or FP more than the other one we change τ.



This leads to another question:

Which value of τ is



Which value of τ is the 'best choice'?

Answer to this question is:

R.O.C. curve

Receiver's Operating Characteristic Curve

→ How to create an ROC curve?

Xi	Yi	Sig(Zi)
X1	1	0.3
X2	1	0.2
X3	1	0.7
X4	0	0.6
X5	0	0.5

step-1 : Sort in descending order of $\sigma(z_i)$

step-2 : For each unique value of $\sigma(z_i)$, take τ & compute $\hat{y}$ (y-pred)

			T= 0.7	T=0.6	T=0.5	T=0.3	T=0.2
Xi	Yi	Sig(Zi)	y_pred	y_pred	y_pred	y_pred	y_pred
X3	1	0.7	1	1	1	1	1
X4	0	0.6	0	1	1	1	1
X5	0	0.5	0	0	1	1	1
X1	1	0.3	0	0	0	1	1
X2	1	0.2	0	0	0	0	1

step-3 : Calculate TPR & FPR for each τ

$\tau = 0.7$

TN 2	FP 0
FN 2	TP 1

$TPR = \frac{1}{3} = 0.3$

$FPR = 0$

$\tau = 0.6$

TN 1	FP 1
FN 2	TP 1

$TPR = \frac{1}{3} = 0.3$

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

$\tau = 0.5$

TN 0	FP 2
FN 2	TP 1

$TPR = \frac{1}{3} = 0.3$

$FPR = 1$

$\tau = 0.3$

TN 0	FP 2
FN 1	TP 2

$TPR = \frac{2}{3} = 0.6$

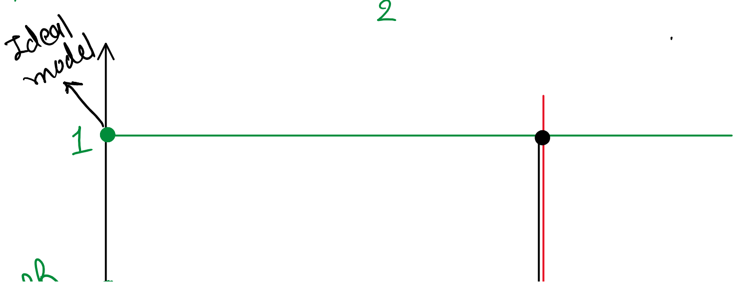
$FPR = 1$

$\tau = 0.2$

TN 0	FP 2
FN 0	TP 3

$TPR = 1$

$FPR = 1$

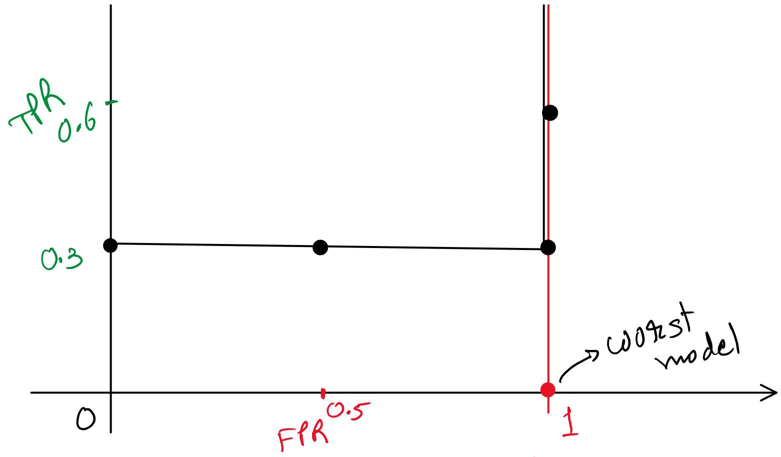


For a worst model:

$TP = 0, TN = 0, FP = p, FN = q$

where $p + q = n$

$TPR = TP/(TP+FN) = 0$



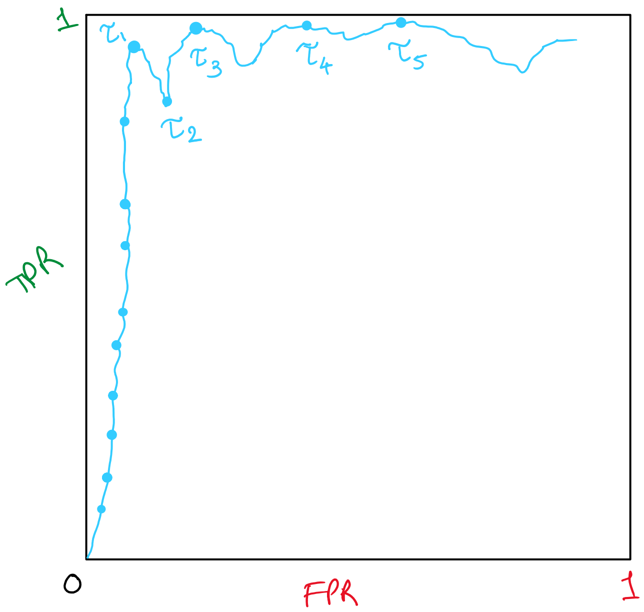
$TPR = TP/(TP+FN) = 0$

$FPR = FP/(FP+TN) = 1$

$TPR = 0$ & $FPR = 1$

For a best model:
$TPR = 1$, $FPR = 0$

A real world ROC curve looks like:



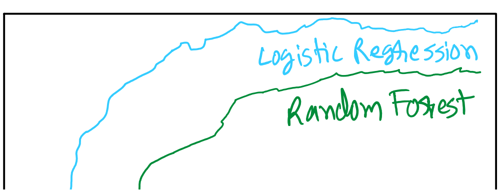
Clearly, τ_1 is better than τ_2. But, which one will you select from τ_1 & τ_3? Usually, classifying points correctly (TPR) is more important than caring about mis-classified points (FPR) (usually, not always) therefore we will select τ_3.

If the difference in TPR is not significant then we choose τ with low FPR. eg, we will choose τ_3 from τ_3, τ_4 & τ_5.

Imagine working with 15-20 confusion matrices to figure out the best choice of τ vs. pointing it out from an ROC curve!

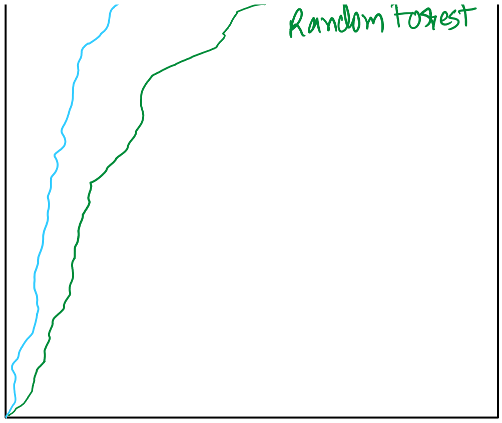
In future, we are going to use multiple different techniques to solve one problem. eg, to classify obese & non-obese people, we are trying out two ML techniques ① Logistic Regression & ② Random Forest

∴ we will get two ROC curves one for each. Let's say they are like below:



Looking at the ROC curves, which technique seems better?

Ans- Logistic Regression

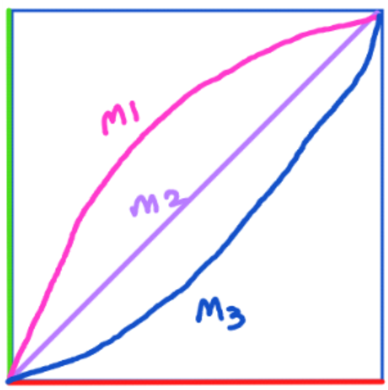


which

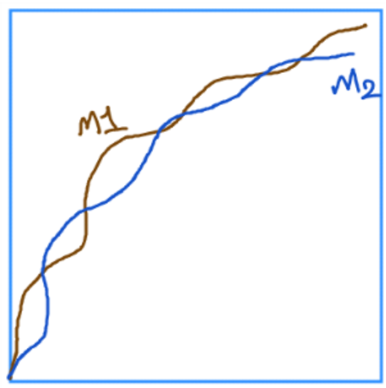
Ans- Logistic Regression

why - Because it has higher TPR compare to Random forest OR it's ROC curve is closer to the ideal model.

Sometimes it is easy to find which method is having higher TPR | closer to the ideal ROC curve :



But sometimes it is not that easy to determine which one is better :



To solve this problem we need **a number** which we can use to compare the methods.

The solution is: Area Under the Curve - AUC

AUC - The method with higher AUC is better than the one with lower AUC.

This AUC is also called: ROC-AUC

The limitation of roc-auc is that it doesn't work well with imballanced data.

For imballanced data, we don't create ROC instead, we create **PRC** (Precision Recall Curve) & use prc-auc

Class Imbalanced

There is no clear distinction available about balanced / imbalanced data. But a widely accepted perception is :

50 - 50 = balanced
60 - 40 = slightly balanced
Imbalance starts → 70 - 30 = slightly imbalanced
80 - 20 = imbalanced
90 - 10
95 - 5
99 - 1
⋮
} Highly imbalanced

→ How do we find weather the data is balanced or not?

① value_counts()

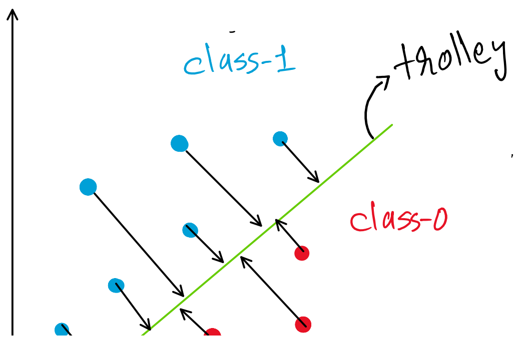
② countplot

→ What are the problems with imbalanced data?

① Accuracy and other a few metrics will no longer be reliable because

② Model starts becoming biased towards the majority class.

Understanding this problem:



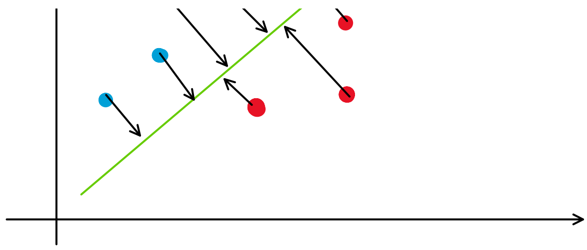
solution-1: class-weights

{● : 1, ● : 2}

$\Sigma(L_i)$

1*

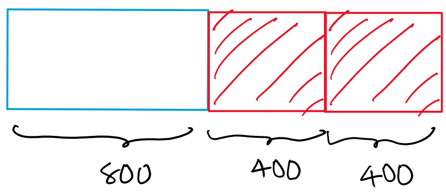
2*



$$\Sigma(L_i)$$

$$-\frac{1}{n}\sum \underbrace{\overset{1*}{y_i}\log\hat{y}_i}_{0} + \underbrace{\overset{2*}{(1-y_i)}\log(1-\hat{y}_i)}_{0}$$

Solution-2 : Over sampling – Duplicate the present minority class data randomly until its size matches with the majority class.



Random sampling with repeatition

Advantages:
① No data loss

Disadvantages:
① Duplication of data
② Chances of overfitting increases.

③ Under sampling – Randomly select as many data points from majority class as in the minority class & only keep them.

eg, if minority class has 100 datapoints & majority class has 900 then we will randomly select 100 data points from majority class & train our model on these 200 datapoints only.

Advantage:
① Computationally in-expensive

Disadvantage:
① Data Loss
② Our randomly selected sample may not capture all the edge cases as well as the complete pattern of the entire data.