

Confusion Matrix

Performance matrix for classification problems.

The rows in a confusion matrix corresponds to prediction &

The cols corresponds to truth (Actual) values.

Actual

<u>Predicted</u>	Has Heart Disease	Does not have heart Disease
	Has Heart Disease	Does not have heart Disease
Has Heart Disease	T.P.	F.P.
Does not have heart disease	F.N.	T.N.

Recall

F.P. → It is an error in data indicates presence of a condition, ~~when~~ such as disease (positive) when in reality it is not present.

False Positive

False Negative

→ It is also an error indicates no presence (result is negative) when in reality it is present.

↓
(less hona chahiye in medical situations)

Precision: Define Out of all samples the classifier labelled as +ve, what fractions were correct.

$$\text{Precision} = \frac{T.P.}{T.P. + F.P.}$$

Recall / Sensitivity / True +ve rate

Out of all +ve samples, what fraction did my classifier pick up correctly.

$$\text{Recall} = \frac{T.P.}{T.P. + F.N.}$$

Specificity (basically, reverse of sensitivity) Out of -ve samples, what fraction did my classifier pick up.

$$\text{Specificity} = \frac{T.N.}{T.N. + F.P.}$$

Sensitivity / TRR
 $\frac{TPR}{FPR} = \frac{TP}{TP + FN}$
TNR
Specificity / TAR
 $\frac{TN}{TN + FP}$
 $\frac{TPR}{FPR} = 1 - Sp$

F1 score = $\frac{2 P.R.}{P + R}$
 $= 2 \times \left(\frac{TP}{TP + FP} \right) \times \left(\frac{TP}{TP + FN} \right)$

$= \frac{2TP^2}{TP^2 + TP \cdot FN + TP^2 + F.P.TP}$
 $= \frac{2TP^2}{2TP^2 + F.P.TP}$

F1 score = $\frac{2TP}{2TP + FP + FN}$

Note
Which one to choose Sensitivity or Specificity??

If correctly identifying +ve is more important than we should choose Sensitivity

If -ve is more important choose Specificity

Note ROC/AUC Curve
Spw (T.P.R vs F.P.R)
False Positive rate = $(1 - \text{Specificity}) = \frac{F.P.}{T.N. + F.P.}$