Metrics to evaluate classification algorithms

18 May 2023 21:51

Different Metrics that we use to evaluate classification algorithms are

1. Confusion Matrix
2. Accuracy
3. Precision
4. Recall / Sensitivity
5. F1 Score.

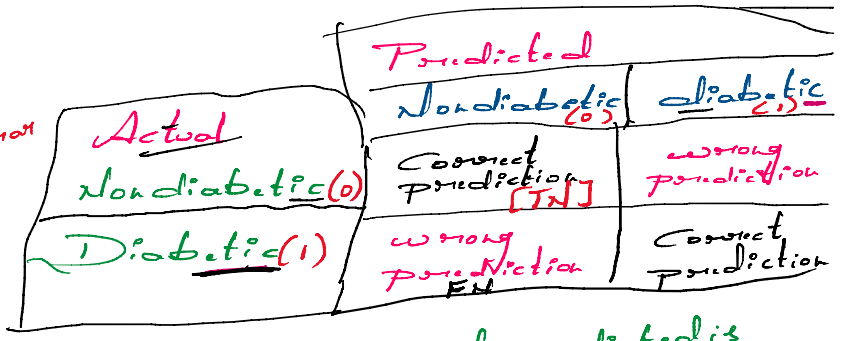
Note
Classification algorithms are evaluated based on misclassifications that model has done.

Confusion Matrix

Actual	Predicted 0	Predicted 1
0	TN	FP
1	FN	TP

Type 1 error →

↳ Type 2 error



actual prediction

- 0 — 0 → TN → Actual is Non diabetic and predicted is non diabetic
- 1 — 1 → TP → Actual is diabetic and predicted prediabetic
- 0 — 1 → FP → Actual is non diabetic and predicted is diabetic
- 1 — 0 → FN → Actual is diabetic and predicted is non diabetic.

Prediction

Actual	Non Cancer (0)	Cancer (1)
Non Cancer (0)	TN	FP
Cancer (1)	FN	TP

Actual	Prediction 0	Prediction 1
0	TN	FP
1	FN	TP

Confusion matrix gives information about no of correct classification and misclassification.

Correct classification = TP + TN

Misclassification = FN + FP

Correct classification + Misclassification = FN + FP

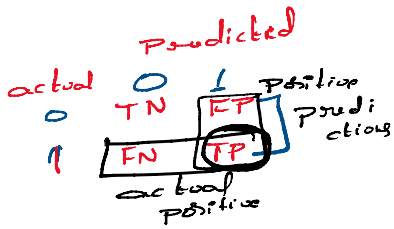
(3) Accuracy Score.

Accuracy : $\frac{\text{Correction predictions}}{\text{Total predictions}}$

Accuracy : $\frac{TP+TN}{TP+TN+FP+FN}$

→ Accuracy measures strength of model.
→ It tells about the percentage of Correct predictions
→ Accuracy score is best metric when data is balanced.
→ If data is not balanced then accuracy is not best metric,
in that case we use recall, precision and F1 score.

(2) Precision



precision : $\frac{TP}{TP+FP}$

out of positive predictions, what percentage are truly positive. [0-1 / 1-1 → how many are 1-1]

(4) Recall

Recall : $\frac{TP}{TP+FN}$

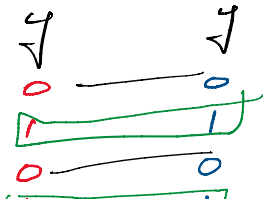
out of all actual positives, what percentage is truly positive. [1-0 / 1-1] → how many are 1-1

(5) F1 score

Harmonic mean of precision and Recall.

F1 score : $\frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$

Eg: actual predicted

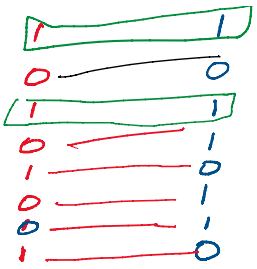


(1) Confusion Matrix

actual \ pred	0	1
0	2	3
1	2	2

TP → 2
TN → 2
FP → 3
FN → 2

[1-1]
[0-0]
[0-1]
[1-0]



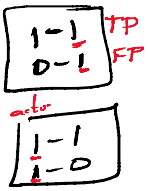
FN → 2

[0 1]
[1 0]

② Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

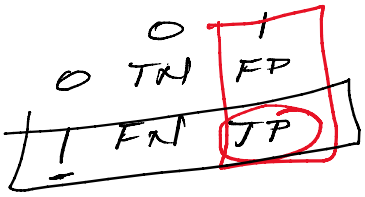
Accuracy = $\frac{2+2}{2+2+3+2} = \frac{4}{9} = 0.66$

③ precision = $\frac{TP}{TP+FP} = \frac{2}{2+3} = \frac{2}{5} = 0.4$



③ Recall = $\frac{TP}{TP+FN} = \frac{2}{2+2} = \frac{2}{4} = 0.5$

④ F_1-Score = $\frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2 \times 0.4 \times 0.5}{0.4+0.5} = 0.44$



Classification

Binary classification
If target column has two categories/output

Eg:- ① Diabetic data ← Diabetic / Non diabetic

Covid19 ← positive / negative

Eg:- Logistic Regression works well with binary classification.

Multi classification
If target column has three or more categories output.

Eg:- ① Iris ← Setosa / Versicolor / Virginica

② Rating ← 1 / 2 / 3 / 4

ROC - AUC Curve

It is one of the metric used to evaluate Binary classification problems such as Logistic regression

ROC → Reciever operator characteristic.

AUC → evaluate over model at different

ROC → Reciever operator characteristic.

using this ROC we evaluate our model at different thresholds

$h_\theta(x) \geq 0.5 \rightarrow y = 0$

$h_\theta(x) \geq 0.5 \rightarrow y = 1$

actual y	$h_\theta(x)$	$\hat{y}$ pred $\hat{y}(0.5)$	$\hat{y}(0.3)$
0	0.2	0	0
1	0.4	0	1
0	0.3	0	1
1	0.7	1	1
1	0.8	1	1
0	0.5	1	1

TP → 1-1

FN → 1-0

TN → 0-0

FP → 0-1

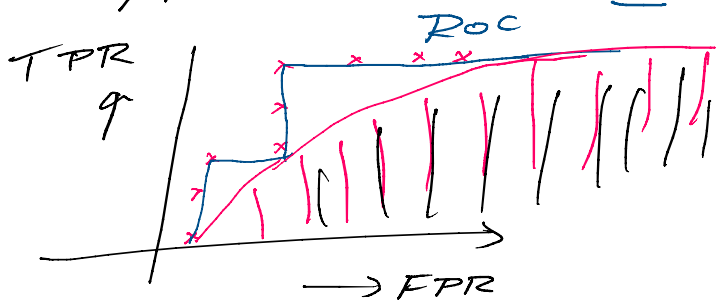
ROC will plot a Graph of TPR v/s FPR.

TPR → True positive rate

$TPR = \frac{TP}{TP+FN} = \frac{2}{2+1} = \frac{2}{3}$ → at 0.5 threshold

FPR → False positive rate

$FPR = \frac{FP}{FP+TN} = \frac{1}{1+2} = \frac{1}{3}$ → at 0.5 threshold



AUC → Area under the under

AUC tells us about strength of the model by measuring area below the ROC curve

* More the area below the ROC curve, better is the model.

Range of AUC = [0, 1]

AUC value near to '0', we say bad model.

AUC value near to '1', we say good model.