09 March 2021 09:06

Confusion Matrix

→ is a mechanism to measure the performance of classifier.
↳ It contains information about actual and predicted classification

Actual class \ Predicted class	Positive	Negative	
Positive	True Positive (TP)	False Negative (FN) Type II errors	Recall (R) or Sensitivity $\frac{TP}{TP+FN}$
Negative	False Positive (FP) Type I errors	True Negative (TN)	Specificity $\frac{TN}{TN+FP}$
	Precision (P) $\frac{TP}{TP+FP}$ +ve predicted value	-ve predictive value $\frac{TN}{TN+FN}$	

Precision or +ve predicted value — The proportion of +ve cases that were correctly identified

-ve predicted value — The proportion of -ve cases that were correctly identified

Sensitivity or Recall — The proportion of actual +ve cases which are correctly identified

Specificity — The proportion of actual -ve cases which are correctly identified

Accuracy - The proportion of the total number of predictions that were correct.

TP → +ve class correctly identified as +ve class
FN → +ve class incorrectly identified as -ve class.
FP → -ve class incorrectly identified as +ve class
TN → -ve class correctly identified as -ve class

Example: Distinguish Spam emails from non-spam emails

→ 100 email examples

Predicted class

→ 100 examples

Actual class \ Predicted class	Spam	Non-Spam
Spam	TP 45	FN 20
Non-Spam	FP 5	TN 30

① Recall (R) $= \frac{TP}{TP+FN} = \frac{45}{45+20} =$ 69.23%

The 69.23% Spam emails are correctly classified and executed from all non-spam emails

② Precision $= \frac{TP}{TP+FP} = \frac{45}{45+5} =$ 90%

The 90% of examples are classified as spam are actually spam

③ Specificity — It is also known True Negative Rate

$= \frac{TN}{TN+FP} = \frac{30}{30+5} =$ 85.71%

The 85.71% non-spam emails are accurately classified and excluded from all spam emails

④ Accuracy

$= \frac{TP+TN}{TP+FP+TN+FN} = \frac{45+30}{45+5+30+20} =$ 75%

The 75% of examples are classified correctly by the classifier.

⑤ F1-Score → is a weighted average of the recall and precision

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

Problem 1

Predicted Class

Actual class		Spam	non-Spam
	Spam	TP = 20	FN = 10
	non-Spam	FP = 30	TN = 40

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

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

Recall = 66.6 %

Precision = 40 %

Accuracy = 60 %

Why higher accuracy does not always imply a good classified.

(6) Classification error = $\frac{\text{error}}{\text{total}} = \frac{FP+FN}{TP+TN+FP+FN}$

(7) Miss classification error : 1 − Accuracy

(8) False +ve Rate or False Alarm Rate = $\frac{FP}{FP+FN}$ → % of −ve we missclassified as +ve

(9) False −ve Rate or Miss Rate = $\frac{FN}{TP+FN}$ → % of +ve we classified as −ve

Problem 2 :

Predicted Class

Actual class		A	B	C
	A	15	2	3
	B	7	15	8
	C	2	3	45

True Positive

Note :

① The total no. of FN → the sum of values in corresponding row (Excluding TP)

② The total no. of FP → The sum of values in the corresponding column (excluding the TP)

③ The total no. of TN → for certain class will be the sum of all columns and rows excluding the class column and row

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \text{Recall} = \frac{TP}{TP+FN}$$

$$P_A = \frac{15}{15+7+2} = \frac{15}{24} = 62.5\% \ (0.625) \qquad R_A = \frac{15}{15+2+3} = \frac{15}{20} = 0.75 = 75\%$$

$$P_B = \frac{15}{15+2+3} = \frac{15}{20} = 75\% \ (0.75) \qquad R_B = \frac{15}{15+7+8} = \frac{15}{30} = 0.5 = 50\%$$

$$P_C = \frac{45}{45+3+8} = \frac{45}{56} = 80\% \ (0.8) \qquad R_C = \frac{45}{45+2+3} = \frac{45}{50} = 0.9 = 90\%$$

✓ Overall Accuracy.

✓ wted avg. precision
✓ wted avg. recall

✓ $A_A =$
$A_B =$
$A_C =$

Programming → Result. Question in