Statistics calculated on confusion matrix

Confusion matrix

Theoretical confusion matrix

		Automatic classification	
		Positive	***Negative***
Manual classification	***Positive***	TP	FN
	Negative	FP	TN

Example

Automatic classification of three groups **A**, **B**, **C**

		Automatic classification		
		A	***B***	***C***
Manual classification	***A***	1	1	1
	B	1	3	0
	C	0	0	3

For group **A**, the matrix can be reduced as:

		Automatic classification	
		Group A	***Not group A***
Manual classification	***Group A***	TP = 1	FN = 2
	Not group A	FP = 1	TN = 6

For group **B**, the matrix can be reduced as:

		Automatic classification	
		Group B	***Not group B***
Manual classification	***Group B***	TP = 3	FN = 1
	Not group B	FP = 1	TN = 5

For group **C**, the matrix can be reduced as:

		Automatic classification	
		Group C	***Not group C***
Manual classification	***Group C***	TP = 3	FN = 0
	Not group C	FP = 1	TN = 6

Explanation of main parameters

TP : True Positive

Number of particles of the group of interest correctly classified.

TN : True Negative

Number of particles of all the other groups classified as other groups.

FP : False Positive

Number of particles of other groups classified in the group of interest.

FN : False Negative

Number of particles of the group of interest classified in the other groups.

General statistics

Accuracy

$$\frac{(TP+TN)}{(TP+TN+FP+FN)}$$

Error

1 – Accuracy

The eight basic ratios

Recall

Also called: Sensitivity, TPR : True Positive Rate, Power, Probability of detection

$\frac{TP}{(TP+FN)}$ = 1 – FNR

Specificity

Also called: TNR :True Negative Rate, Selectivity

$\frac{TN}{(TN+FP)}$ = 1 – FPR

Precision

Also called: PPV : Positive Predicted Value, Reproducibility, Repeatability

$\frac{TP}{(TP+FP)}$ = 1 – FDR

NPV : Negative Predicted Value

$\frac{TN}{(TN+FN)}$ = 1 – FOR

FPR : False Positive Rate

Also called: alpha, Type I Error, p-Value

$\frac{FP}{(FP+TN)}$ = 1 – Specificity

FNR : False Negative Rate

Also called: beta, Type II Error

$\frac{FN}{(TP+FN)}$ = 1 – Recall

FDR : False Discovery Rate

Also called : q-Value

$\frac{FP}{(TP+FP)}$ = 1 – Precision

FOR : False Omission Rate

$$\frac{FN}{(FN+TN)} = 1 - \text{NPV}$$

The four ratios of ratios

LRPT : Likelihood Ratio for Positive Tests

$$\frac{TP/(TP+FN)}{FP/(FP+TN)} = \frac{Recall}{(1-Specificity)} = \frac{Recall}{FPR}$$

LRNT : Likelihood Ratio for Negative Tests

$$\frac{FN/(FN+TP)}{TN/(TN+FP)} = \frac{(1-Recall)}{Specificity} = \frac{FNR}{Specificity}$$

LRPS : Likelihood Ratio for Positive Subjects

$$\frac{TP/(TP+FP)}{FN/(FN+TN)} = \frac{Precision}{(1-NPV)} = \frac{Precision}{FOR}$$

LRNS : Likelihood Ratio for Negative Subjects

$$\frac{FP/(FP+TP)}{TN/(TN+FN)} = \frac{(1-Precision)}{(1-FOR)} = \frac{FDR}{NPV}$$

Additional statistics

Fmes : F-measure

Also called: F1-score, harmonic mean of precision and recall

$$2*(\frac{(Precision*Recall)}{(Precision+Recall)})$$

BalAcc : Balanced accuracy

$$\frac{(Specificity+Recall)}{2}$$

MCC : Matthews Correlation Coefficient

Attention: If any sum of the denominator is **0**, the total denominator can be set to **1**

$$\frac{(TP*TN)-(FP*FN)}{\sqrt{((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))}}$$

Chisq : χ^2

Also called : Significance

$$\frac{[(TP*TN)-(FP*FN)]^2*(TP+TN+FP+FN)}{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}$$

Auto_Manu : Difference between Automatic and Manual classification

$$(TP+FP)-(TP+FN)$$

Dissimilarity Index of Bray Curtis

$$\frac{|(Auto\_Manu)|}{\sum(TP+FP)+\sum(TP+FN)}$$