

Confusion matrix

From Wikipedia, the free encyclopedia

In the field of machine learning and specifically the problem of statistical classification, a **confusion matrix**, also known as an error matrix,^[4] is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a **matching matrix**). Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class (or vice versa).^[2] The name stems from the fact that it makes it easy to see if the system is confusing two classes (i.e. commonly mislabelling one as another).

It is a special kind of contingency table, with two dimensions ("actual" and "predicted"), and identical sets of "classes" in both dimensions (each combination of dimension and class is a variable in the contingency table).

Contents

- 1 Example
- 2 Table of confusion
- 3 References
- 4 External links

Example

If a classification system has been trained to distinguish between cats, dogs and rabbits, a confusion matrix will summarize the results of testing the algorithm for further inspection. Assuming a sample of 27 animals — 8 cats, 6 dogs, and 13 rabbits, the resulting confusion matrix could look like the table below:

		Predicted		
		Cat	Dog	Rabbit
Actual class	Cat	5	3	0
	Dog	2	3	1
	Rabbit	0	2	11

In this confusion matrix, of the 8 actual cats, the system predicted that three were dogs, and of the six dogs, it predicted that one was a rabbit and two were cats. We can see from the matrix that the system in question has trouble distinguishing between cats and dogs, but can make the distinction between rabbits and other types of animals pretty well. All correct guesses are located in the diagonal of the table, so it's easy to visually inspect the table for errors, as they will be represented by values outside the diagonal.

Table of confusion

In predictive analytics, a **table of confusion** (sometimes also called a **confusion matrix**), is a table with two rows and two columns that reports the number of *false positives*, *false negatives*, *true positives*, and *true negatives*. This allows more detailed analysis than mere proportion of correct guesses (accuracy). Accuracy is not a reliable metric for the real performance of a classifier, because it will yield misleading results if the data set is unbalanced (that is, when the number of samples in different classes vary greatly). For example, if there were 95 cats and only 5 dogs in the data set, the classifier could easily be biased into classifying all the samples as cats. The overall accuracy would be 95%, but in practice the classifier would have a 100% recognition rate for the cat class but a 0% recognition rate for the dog class.

Assuming the confusion matrix above, its corresponding table of confusion, for the cat class, would be:

5 true positives (actual cats that were correctly classified as cats)	3 false negatives (cats that were incorrectly marked as dogs)
2 false positives (dogs that were incorrectly labeled as cats)	17 true negatives (all the remaining animals, correctly classified as non-cats)

The final table of confusion would contain the average values for all classes combined.

Let us define an experiment from **P** positive instances and **N** negative instances for some condition. The four outcomes can be formulated in a 2x2 *confusion matrix*, as follows:

Terminology and derivations from a confusion matrix

true positive (TP)
eqv. with hit

true negative (TN)
eqv. with correct rejection

false positive (FP)
eqv. with false alarm, Type I error

false negative (FN)
eqv. with miss, Type II error

sensitivity or true positive rate (TPR)
eqv. with hit rate, recall

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

specificity (SPC) or true negative rate (TNR)
$$SPC = \frac{TN}{N} = \frac{TN}{FP + TN}$$

precision or positive predictive value (PPV)
$$PPV = \frac{TP}{TP + FP}$$

recall (recall)
$$recall = \frac{TP}{TP + FN}$$

negative predictive value (NPV)
$$NPV = \frac{TN}{TN + FN}$$

fall-out or false positive rate (FPR)
$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - SPC$$

false discovery rate (FDR)
$$FDR = \frac{FP}{FP + TP} = 1 - PPV$$

miss rate or false negative rate (FNR)
$$FNR = \frac{FN}{P} = \frac{FN}{FN + TP} = 1 - TPR$$

accuracy (ACC)
$$ACC = \frac{TP + TN}{P + N}$$

F1 score
is the harmonic mean of precision and sensitivity
$$F1 = \frac{2TP}{2TP + FP + FN}$$

Matthews correlation coefficient (MCC)
$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Informedness = Sensitivity + Specificity - 1

Markedness = Precision + NPV - 1

Sources: Fawcett (2006), Powers (2011), and Ting (2011) ^[1] ^[2] ^[3]

		Predicted condition			
Total population		Predicted Condition positive	Predicted Condition negative	Prevalence $= \frac{\Sigma \text{Condition positive}}{\Sigma \text{Total population}}$	
True condition	condition positive	True positive	False Negative (Type II error)	True positive rate (TPR), Sensitivity, Recall $= \frac{\Sigma \text{True positive}}{\Sigma \text{Condition positive}}$	False negative rate (FN Miss rate $= \frac{\Sigma \text{False negative}}{\Sigma \text{Condition positive}}$
	condition negative	False Positive (Type I error)	True negative	False positive rate (FPR), Fall-out $= \frac{\Sigma \text{False positive}}{\Sigma \text{Condition negative}}$	True negative rate (TN Specificity (SPC) $= \frac{\Sigma \text{True negative}}{\Sigma \text{Condition negative}}$
Accuracy (ACC) = $\frac{\Sigma \text{True positive} + \Sigma \text{True negative}}{\Sigma \text{Total population}}$		Positive predictive value (PPV), Precision $= \frac{\Sigma \text{True positive}}{\Sigma \text{Test outcome positive}}$	False omission rate (FOR) $= \frac{\Sigma \text{False negative}}{\Sigma \text{Test outcome negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR-}}$
		False discovery rate (FDR) $= \frac{\Sigma \text{False positive}}{\Sigma \text{Test outcome positive}}$	Negative predictive value (NPV) $= \frac{\Sigma \text{True negative}}{\Sigma \text{Test outcome negative}}$	Negative likelihood ratio (LR-) = $\frac{\text{FNR}}{\text{TNR}}$	

References

1. Fawcett, Tom (2006). "An Introduction to ROC Analysis". *Pattern Recognition Letters*. **27** (8): 861 – 874. doi:10.1016/j.patrec.2005.10.010.
2. Powers, David M W (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" (PDF). *Journal of Machine Learning Technologies*. **2** (1): 37–63.
3. Ting, Kai Ming (2011). *Encyclopedia of machine learning*. Springer. ISBN 978-0-387-30164-8.
4. Stehman, Stephen V. (1997). "Selecting and interpreting measures of thematic classification accuracy". *Remote Sensing of Environment*. **62** (1): 77–89. doi:10.1016/S0034-4257(97)00083-7.

External links

- Theory about the confusion matrix (http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html)
- GM-RKB Confusion Matrix concept page (http://www.gabormelli.com/RKB/Confusion_Matrix)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Confusion_matrix&oldid=743839635"

Categories: Machine learning | Statistical classification

- This page was last modified on 11 October 2016, at 15:25.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.