

Title: Confusion matrix

URL: [https://en.wikipedia.org/wiki/Confusion\\_matrix](https://en.wikipedia.org/wiki/Confusion_matrix)

PageID: 847558

Categories: Category:Machine learning, Category:Statistical classification

Source: Wikipedia (CC BY-SA 4.0). Content may require attribution.

-----

In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as error matrix, 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 row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class, or vice versa – both variants are found in the literature. The diagonal of the matrix therefore represents all instances that are correctly predicted. The name stems from the fact that it makes it easy to see whether the system is confusing two classes (i.e. commonly mislabeling 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).

#### Example

Given a sample of 12 individuals, 8 that have been diagnosed with cancer and 4 that are cancer-free, where individuals with cancer belong to class 1 (positive) and non-cancer individuals belong to class 0 (negative), we can display that data as follows:

Assume that we have a classifier that distinguishes between individuals with and without cancer in some way, we can take the 12 individuals and run them through the classifier. The classifier then makes 9 accurate predictions and misses 3: 2 individuals with cancer wrongly predicted as being cancer-free (sample 1 and 2), and 1 person without cancer that is wrongly predicted to have cancer (sample 9).

Notice, that if we compare the actual classification set to the predicted classification set, there are 4 different outcomes that could result in any particular column. One, if the actual classification is positive and the predicted classification is positive (1,1), this is called a true positive result because the positive sample was correctly identified by the classifier. Two, if the actual classification is positive and the predicted classification is negative (1,0), this is called a false negative result because the positive sample is incorrectly identified by the classifier as being negative. Third, if the actual classification is negative and the predicted classification is positive (0,1), this is called a false positive result because the negative sample is incorrectly identified by the classifier as being positive. Fourth, if the actual classification is negative and the predicted classification is negative (0,0), this is called a true negative result because the negative sample gets correctly identified by the classifier.

We can then perform the comparison between actual and predicted classifications and add this information to the table, making correct results appear in green so they are more easily identifiable.

The template for any binary confusion matrix uses the four kinds of results discussed above (true positives, false negatives, false positives, and true negatives) along with the positive and negative classifications. The four outcomes can be formulated in a 2x2 confusion matrix, as follows:

The color convention of the three data tables above were picked to match this confusion matrix, in order to easily differentiate the data.

Now, we can simply total up each type of result, substitute into the template, and create a confusion matrix that will concisely summarize the results of testing the classifier:

$$8 + 4 = 12$$

In this confusion matrix, of the 8 samples with cancer, the system judged that 2 were cancer-free, and of the 4 samples without cancer, it predicted that 1 did have cancer. All correct predictions are located in the diagonal of the table (highlighted in green), so it is easy to visually inspect the table for prediction errors, as values outside the diagonal will represent them. By summing up the 2 rows of the confusion matrix, one can also deduce the total number of positive (P) and negative (N) samples in the original dataset, i.e.  $P = TP + FN$  and  $N = FP + TN$ .

#### 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 true positives, false negatives, false positives, and true negatives. This allows more detailed analysis than simply observing the proportion of correct classifications (accuracy). Accuracy will yield misleading results if the data set is unbalanced; that is, when the numbers of observations in different classes vary greatly.

For example, if there were 95 cancer samples and only 5 non-cancer samples in the data, a particular classifier might classify all the observations as having cancer. The overall accuracy would be 95%, but in more detail the classifier would have a 100% recognition rate (sensitivity) for the cancer class but a 0% recognition rate for the non-cancer class. F1 score is even more unreliable in such cases, and here would yield over 97.4%, whereas informedness removes such bias and yields 0 as the probability of an informed decision for any form of guessing (here always guessing cancer).

According to Davide Chicco and Giuseppe Jurman, the most informative metric to evaluate a confusion matrix is the Matthews correlation coefficient (MCC).

Other metrics can be included in a confusion matrix, each of them having their significance and use.

[view](#)

[talk](#)

[edit](#)

Some researchers have argued that the confusion matrix, and the metrics derived from it, do not truly reflect a model's knowledge. In particular, the confusion matrix cannot show whether correct predictions were reached through sound reasoning or merely by chance (a problem known in philosophy as epistemic luck). It also does not capture situations where the facts used to make a prediction later change or turn out to be wrong (defeasibility). This means that while the confusion matrix is a useful tool for measuring classification performance, it may give an incomplete picture of a model's true reliability.

#### Confusion matrices with more than two categories

Confusion matrix is not limited to binary classification and can be used in multi-class classifiers as well. The confusion matrices discussed above have only two conditions: positive and negative. For example, the table below summarizes communication of a whistled language between two speakers, with zero values omitted for clarity.

#### Confusion matrices in multi-label and soft-label classification

Confusion matrices are not limited to single-label classification (where only one class is present) or hard-label settings (where classes are either fully present, 1, or absent, 0). They can also be extended to Multi-label classification (where multiple classes can be predicted at once) and soft-label classification (where classes can be partially present).

One such extension is the Transport-based Confusion Matrix (TCM), which builds on the theory of optimal transport and the principle of maximum entropy. TCM applies to single-label, multi-label, and soft-label settings. It retains the familiar structure of the standard confusion matrix: a square matrix sized by the number of classes, with diagonal entries indicating correct predictions and off-diagonal entries indicating confusion. In the single-label case, TCM is identical to the standard confusion matrix.

TCM follows the same reasoning as the standard confusion matrix: if class A is overestimated (its predicted value is greater than its label value) and class B is underestimated (its predicted value is less than its label value), A is considered confused with B, and the entry (B, A) is increased. If a class is both predicted and present, it is correctly identified, and the diagonal entry (A, A) increases. Optimal transport and maximum entropy are used to determine the extent to which these entries are updated.

TCM enables clearer comparison between predictions and labels in complex classification tasks, while maintaining a consistent matrix format across settings.

See also

Positive and negative predictive values

References

v

t

e

Alternant

Anti-diagonal

Anti-Hermitian

Anti-symmetric

Arrowhead

Band

Bidiagonal

Bisymmetric

Block-diagonal

Block

Block tridiagonal

Boolean

Cauchy

Centrosymmetric

Conference

Complex Hadamard

Copositive

Diagonally dominant

Diagonal

Discrete Fourier Transform

Elementary

Equivalent

Frobenius

Generalized permutation

Hadamard

Hankel

Hermitian  
Hessenberg  
Hollow  
Integer  
Logical  
Matrix unit  
Metzler  
Moore  
Nonnegative  
Pentadiagonal  
Permutation  
Persymmetric  
Polynomial  
Quaternionic  
Signature  
Skew-Hermitian  
Skew-symmetric  
Skyline  
Sparse  
Sylvester  
Symmetric  
Toeplitz  
Triangular  
Tridiagonal  
Vandermonde  
Walsh  
Z  
Exchange  
Hilbert  
Identity  
Lehmer  
Of ones  
Pascal  
Pauli  
Redheffer  
Shift  
Zero  
Companion

Convergent  
Defective  
Definite  
Diagonalizable  
Hurwitz-stable  
Positive-definite  
Stieltjes  
Congruent  
Idempotent or Projection  
Invertible  
Involutory  
Nilpotent  
Normal  
Orthogonal  
Unimodular  
Unipotent  
Unitary  
Totally unimodular  
Weighing  
Adjugate  
Alternating sign  
Augmented  
Bézout  
Carleman  
Cartan  
Circulant  
Cofactor  
Commutation  
Confusion  
Coxeter  
Distance  
Duplication and elimination  
Euclidean distance  
Fundamental (linear differential equation)  
Generator  
Gram  
Hessian  
Householder

Jacobian  
Moment  
Payoff  
Pick  
Random  
Rotation  
Routh-Hurwitz  
Seifert  
Shear  
Similarity  
Symplectic  
Totally positive  
Transformation  
Centering  
Correlation  
Covariance  
Design  
Doubly stochastic  
Fisher information  
Hat  
Precision  
Stochastic  
Transition  
Adjacency  
Biadjacency  
Degree  
Edmonds  
Incidence  
Laplacian  
Seidel adjacency  
Tutte  
Cabibbo–Kobayashi–Maskawa  
Density  
Fundamental (computer vision)  
Fuzzy associative  
Gamma  
Gell-Mann  
Hamiltonian

Irregular

Overlap

S

State transition

Substitution

Z (chemistry)

Jordan normal form

Linear independence

Matrix exponential

Matrix representation of conic sections

Perfect matrix

Pseudoinverse

Row echelon form

Wronskian

Mathematics portal

List of matrices

Category:Matrices (mathematics)