Search Wikipedia

Q



Main page Contents Current events Random article

**About Wikipedia** Contact us Donate Contribute

Help Learn to edit Community portal Recent changes Upload file

Tools What links here Related changes Special pages Permanent link Page information Cite this page Wikidata item

Print/export Download as PDF Printable version

Languages العربية Deutsch Español Français

한국어

Italiano

日本語

Português

0

文 5 more

Edit links

#### Confusion matrix From Wikipedia, the free encyclopedia

Article Talk

In the field of machine learning and specifically the problem of statistical classification, a **confusion** matrix, also known as an error matrix, [10] 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. [11] 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). **Contents** [hide]

## 1 Example 2 Table of confusion 3 Confusion matrices with more than two categories 4 See also 5 References

Given a sample of 12 individuals, 8 that have been diagnosed with cancer and 4 that are cancer-free,

**Individual Number** 

**Actual Classification** 

**Predicted Classification** 

Result

Example [edit]

(negative), we can display that data as follows: 1 2 3 4 5 6 7 8 9 10 11 12 **Individual Number** 

where individuals with cancer belong to class 1 (positive) and non-cancer individuals belong to class 0

Actual Classification | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 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

Actual Classification | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0

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). **Individual Number** 1 2 3 4 5 6 7 8 9 10 11 12

Predicted Classification 0 0 1 1 1 1 1 1 1 0 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.

10

11 | 12

Sources: [12][13][14][15][16][17][18][19]

to match this confusion matrix, in order

view ·talk ·edit

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 2×2 *confusion matrix*, as follows:

FN FN TP TP TP TP TP FP TN TN TN

		Predicted condition						
	Total population = P + N	Positive (PP)	Negative (PN)					
Actual condition	Positive (P)	True positive (TP)	False negative (FN)					
	Negative (N)	False positive (FP)	True negative (TN)					
The color convention of the three data tables above were picked to easily differentiate the data.								

matrix that will concisely summarize the results of testing the classifier:

**Predicted condition** Total Cancer Non-cancer

Now, we can simply total up each type of result, substitute into the template, and create a confusion

8 + 4 = 125

### condition Cancer 6 2 Non-cancer Actual 3 In this confusion matrix, of the 8 samples with cancer, the system judged that 2 were cancer-free, and of number of positive (P) and negative (N) samples in the original dataset, i.e. P = TP + FN and N = FP + TN.

Total population

# from a confusion matrix

**Terminology and derivations** 

condition positive (P)

Edit | View history

Read

the number of real positive cases in the data condition negative (N) the number of real negative cases in the data

true positive (TP)

A test result that correctly indicates the presence of a condition or characteristic

true negative (TN) A test result that correctly indicates the absence of a condition or

characteristic false positive (FP) A test result which wrongly indicates that a particular condition or

attribute is present false negative (FN) A test result which wrongly indicates that a particular condition or

attribute is absent sensitivity, recall, hit rate, or true positive rate (TPR)  $TPR = \frac{TP}{P} = \frac{TP}{TP + FN} = 1 - FNR$ 

specificity, selectivity or true negative rate (TNR)  $ext{TNR} = rac{ ext{TN}}{ ext{N}} = rac{ ext{TN}}{ ext{TN} + ext{FP}} = 1 - ext{FPR}$ precision or positive predictive value (PPV)  $PPV = \frac{TP}{TP + FP} = 1 - FDR$ 

negative predictive value (NPV)  $NPV = \frac{TN}{TN + FN} = 1 - FOR$ 

miss rate or false negative rate (FNR)  $FNR = \frac{FN}{P} = \frac{FN}{FN + TP} = 1 - TPR$ fall-out or false positive rate (FPR)  $ext{FPR} = rac{ ext{FP}}{ ext{N}} = rac{ ext{FP}}{ ext{FP} + ext{TN}} = 1 - ext{TNR}$ false discovery rate (FDR)  $FDR = \frac{FP}{FP + TP} = 1 - PPV$ false omission rate (FOR)

 $FOR = \frac{FN}{FN + TN} = 1 - NPV$ Positive likelihood ratio (LR+)  $LR + = \frac{TPR}{FPR}$ **Negative likelihood ratio (LR-)**  $LR-=rac{FNR}{TNR}$ 

prevalence threshold (PT)  $\mathrm{PT} = rac{\sqrt{\mathrm{FPR}}}{\sqrt{\mathrm{TPR}} + \sqrt{\mathrm{FPR}}}$ threat score (TS) or critical success index (CSI)  $TS = \frac{}{TP + FN + FP}$ **Prevalence** 

 $\overline{P+N}$ accuracy (ACC)  $ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN}$ balanced accuracy (BA)  $BA = \frac{TPR + TNR}{2}$ F1 score is the harmonic mean of precision and sensitivity:  $ext{F}_1 = 2 imes rac{ ext{PPV} imes ext{TPR}}{ ext{PPV} + ext{TPR}} = rac{2 ext{TP}}{2 ext{TP} + ext{FP} + ext{FN}}$ phi coefficient ( $\phi$  or  $r_{\phi}$ ) or Matthews correlation coefficient (MCC)  $MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$ Fowlkes-Mallows index (FM)  $\frac{\overline{TP}}{\overline{P} + \overline{P}N} = \sqrt{PPV \times TPR}$  $\mathrm{FM} = \sqrt{\frac{11}{TP + FP}} imes \frac{11}{TP + FN}$ 

informedness or bookmaker informedness (BM) BM = TPR + TNR - 1markedness (MK) or deltaP (Δp) MK = PPV + NPV - 1Diagnostic odds ratio (DOR)  $DOR = \frac{LR+}{LR-}$ Sources: Fawcett (2006),<sup>[1]</sup> Piryonesi and El-Diraby (2020),<sup>[2]</sup> Powers (2011),<sup>[3]</sup> Ting (2011),<sup>[4]</sup> CAWCR,<sup>[5]</sup> D. Chicco & G. Jurman (2020, 2021),<sup>[6][7]</sup> Tharwat (2018).<sup>[8]</sup> Balayla (2020)<sup>[9]</sup> 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

Sources: [21][22][23][24][25][26][27][28][29] view talk edit

Prevalence threshold (PT)

Table of confusion [edit]

misleading results if the data set is unbalanced; that is, when the numbers of observations in different classes vary greatly.

**Predicted condition** 

hit

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

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.

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

Informedness, bookmaker informedness (BM)

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

Positive (PP) **Negative (PN)**  $= \frac{\sqrt{\mathsf{TPR} {\times} \mathsf{FPR}} {-} \mathsf{FPR}}{\mathsf{TPR} {-} \mathsf{FPR}}$ = P + N= TPR + TNR - 1 True positive rate (TPR), recall, sensitivity False negative rate (FNR), False negative (FN), True positive (TP), Positive (P) (SEN), probability of detection, hit rate, power miss rate type II error, miss,

Actual condition  $=\frac{FN}{P}=1-TPR$  $=\frac{TP}{P}=1-FNR$ underestimation False positive rate (FPR), True negative rate (TNR), False positive (FP), True negative (TN), specificity (SPC), selectivity **Negative (N)** probability of false alarm, fall-out type I error, false alarm, correct rejection  $=\frac{FP}{N}=1-TNR$  $=\frac{TN}{N}=1-FPR$ overestimation Positive predictive value (PPV), False omission rate Positive likelihood ratio (LR+) Negative likelihood ratio (LR-) Prevalence (FOR)  $=\frac{TPR}{FPR}$  $=\frac{P}{P+N}$  $=\frac{TP}{PP}=1-FDR$  $=\frac{FN}{PN}=1-NPV$ False discovery rate (FDR) Negative predictive value Accuracy (ACC) Markedness (MK), deltaP (Δp) Diagnostic odds ratio (DOR) =  $\frac{LR+}{LR-}$  $=\frac{TP+TN}{P+N}$  $=\frac{FP}{PP}=1-PPV$  $(NPV) = \frac{TN}{PN} = 1 - FOR$ = PPV + NPV - 1Matthews correlation coefficient (MCC) Balanced accuracy Threat score (TS), critical success index Fowlkes-Mallows index  $=\sqrt{\mathsf{TPR}\times\mathsf{TNR}\times\mathsf{PPV}\times\mathsf{NPV}}$  $= \frac{2PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$  $(BA) = \frac{TPR + TNR}{2}$ (CSI), Jaccard index =  $\frac{TP}{TP + FN + FP}$  $(FM) = \sqrt{PPV \times TPR}$  $-\sqrt{\mathsf{FNR}\times\mathsf{FPR}\times\mathsf{FOR}\times\mathsf{FDR}}$ Confusion matrices with more than two categories [edit] Confusion matrix is not limited to binary classification and can be used in multi-class classifiers as well. [30] The confusion matrices discussed above have only two conditions:

Vowel produced 15

u

0

a

е

е

positive and negative. For example, the table below summarizes communication of a whistled language between two speakers, zero values omitted for clarity. [31]

a			79	5						
0			4	15	3					
u				2	2					
See also [edit]										
Positive and negative predictive values										
References [edit]										
1. ^ Fawcett, Tom (2006). "An Introduction to ROC Analysis" (PDF). Pattern										

#### Recognition Letters. 27 (8): 861–874. doi:10.1016/j.patrec.2005.10.010 ℃. 2. A Pirvonesi S. Madeh; El-Diraby Tamer E. (2020-03-01). "Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index". Journal of

555X.0000512℃.

Perceived

vowel

3. ^ Powers, David M. W. (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" ... Journal of Machine Learning Technologies. 2 (1): 37-63.

Infrastructure Systems. 26 (1): 04019036. doi:10.1061/(ASCE)IS.1943-

4. ^ Ting, Kai Ming (2011). Sammut, Claude; Webb, Geoffrey I. (eds.). Encyclopedia of machine learning. Springer. doi:10.1007/978-0-387-30164-8 ₺. ISBN 978-0-387-30164-5. A Brooks, Harold; Brown, Barb; Ebert, Beth; Ferro, Chris; Jolliffe, Ian; Koh, Tieh-Yong; Roebber, Paul; Stephenson, David (2015-01-26). "WWRP/WGNE Joint Working Group" on Forecast Verification Research" . Collaboration for Australian Weather and Climate

Research. World Meteorological Organisation. Retrieved 2019-07-17.

BMC Genomics. 21 (1): 6-1-6-13. doi:10.1186/s12864-019-6413-7 △. PMC 6941312 PMID 31898477 ℃. 7. ^ Chicco D.; Toetsch N.; Jurman G. (February 2021). "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness,

1-22. doi:10.1186/s13040-021-00244-z2. PMC 7863449 . PMID 335414102.

8. \* Tharwat A. (August 2018). "Classification assessment methods" . Applied

classification accuracy". Remote Sensing of Environment. 62 (1): 77–89.

Recognition Letters. 27 (8): 861–874. Bibcode:2006PaReL..27..861F 2.

doi:10.1016/j.patrec.2005.10.010 2. S2CID 2027090 2.

555X.0000512 ℃. S2CID 213782055 ℃.

V •T •E

Computing and Informatics. doi:10.1016/j.aci.2018.08.003

and markedness in two-class confusion matrix evaluation" ∠. BioData Mining. 14 (13):

6. A Chicco D.; Jurman G. (January 2020). "The advantages of the Matthews correlation

- 9. A Balayla, Jacques (2020). "Prevalence threshold (φe) and the geometry of screening curves" . PLoS One. 15 (10). doi:10.1371/journal.pone.0240215 . 10. A Stehman, Stephen V. (1997). "Selecting and interpreting measures of thematic
- Bibcode:1997RSEnv..62...77S \(\mathbb{Z}\). doi:10.1016/S0034-4257(97)00083-7 \(\mathbb{Z}\). 11. ^ Powers, David M. W. (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" ... Journal of Machine Learning Technologies. 2 (1): 37–63. S2CID 55767944 €. 12. \* Fawcett, Tom (2006). "An Introduction to ROC Analysis" in (PDF). Pattern
- 13. A Piryonesi S. Madeh; El-Diraby Tamer E. (2020-03-01). "Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index". Journal of Infrastructure Systems. 26 (1): 04019036. doi:10.1061/(ASCE)IS.1943-
- 14. ^ Powers, David M. W. (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" ... Journal of Machine Learning Technologies. 2 (1): 37–63.
- machine learning. Springer. doi:10.1007/978-0-387-30164-8 ₺. ISBN 978-0-387-30164-16. A Brooks, Harold; Brown, Barb; Ebert, Beth; Ferro, Chris; Jolliffe, Ian; Koh, Tieh-Yong; Roebber, Paul; Stephenson, David (2015-01-26). "WWRP/WGNE Joint Working Group on Forecast Verification Research" . Collaboration for Australian Weather and Climate Research. World Meteorological Organisation. Retrieved 2019-07-17.

15. ^ Ting, Kai Ming (2011). Sammut, Claude; Webb, Geoffrey I. (eds.). Encyclopedia of

Computing and Informatics. 17: 168–192. doi:10.1016/j.aci.2018.08.003 20. ^ Chicco D., Jurman G. (January 2020). "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation" 2. BMC Genomics. 21 (1): 6-1–6-13. doi:10.1186/s12864-019-6413-7 ₺. PMC 6941312 PMID 31898477 ℃. 21. A Balayla, Jacques (2020). "Prevalence threshold (\$\phi\$e) and the geometry of screening

19. ^ Tharwat A. (August 2018). "Classification assessment methods" ☑. Applied

18. ^ Chicco D, Toetsch N, Jurman G (February 2021). "The Matthews correlation

13. doi:10.1186/s13040-021-00244-z ₽. PMC 7863449 . PMID 33541410 ₽.

17. ^ Chicco D, Jurman G (January 2020). "The advantages of the Matthews correlation

PMID 31898477 2.

555X.0000512 ℃.

coefficient (MCC) over F1 score and accuracy in binary classification evaluation" 2.

BMC Genomics. 21 (1): 6-1–6-13. doi:10.1186/s12864-019-6413-7 △. PMC 6941312

coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness,

and markedness in two-class confusion matrix evaluation" \alpha. BioData Mining. 14 (13):

curves" ∠. PLoS One. **15** (10). doi:10.1371/journal.pone.0240215 ∠. 22. \* Fawcett, Tom (2006). "An Introduction to ROC Analysis" (PDF). Pattern Recognition Letters. 27 (8): 861–874. doi:10.1016/j.patrec.2005.10.010 ℃. 23. ^ Piryonesi S. Madeh; El-Diraby Tamer E. (2020-03-01). "Data Analytics in Asset

Management: Cost-Effective Prediction of the Pavement Condition Index". Journal of

24. ^ Powers, David M. W. (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" ... Journal of Machine Learning Technologies. 2 (1): 37-63.

Infrastructure Systems. 26 (1): 04019036. doi:10.1061/(ASCE)IS.1943-

26. A Brooks, Harold; Brown, Barb; Ebert, Beth; Ferro, Chris; Jolliffe, Ian; Koh, Tieh-Yong; Roebber, Paul; Stephenson, David (2015-01-26). "WWRP/WGNE Joint Working Group on Forecast Verification Research" 2. Collaboration for Australian Weather and Climate Research. World Meteorological Organisation. Retrieved 2019-07-17.

25. ^ Ting, Kai Ming (2011). Sammut, Claude; Webb, Geoffrey I. (eds.). Encyclopedia of

machine learning. Springer. doi:10.1007/978-0-387-30164-8 ☑. ISBN 978-0-387-30164-

coefficient (MCC) over F1 score and accuracy in binary classification evaluation" ≥. BMC Genomics. 21 (1): 6-1–6-13. doi:10.1186/s12864-019-6413-7 △. PMC 6941312 PMID 31898477 ℃. 28. ^ Chicco D, Toetsch N, Jurman G (February 2021). "The Matthews correlation

27. ^ Chicco D, Jurman G (January 2020). "The advantages of the Matthews correlation

- coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation" . BioData Mining. 14 (13): 1-22. doi:10.1186/s13040-021-00244-z2. PMC 7863449 . PMID 335414102.
- Computing and Informatics. doi:10.1016/j.aci.2018.08.003 30. A Piryonesi S. Madeh; El-Diraby Tamer E. (2020-03-01). "Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index". Journal of

29. ^ Tharwat A. (August 2018). "Classification assessment methods" ∠. Applied

Infrastructure Systems. 26 (1): 04019036. doi:10.1061/(ASCE)IS.1943-

555X.0000512 ℃. S2CID 213782055 ℃. 31. A Rialland, Annie (August 2005). "Phonological and phonetic aspects of whistled languages". Phonology. 22 (2): 237–271. CiteSeerX 10.1.1.484.4384 doi:10.1017/S0952675705000552 2. S2CID 18615779 2.

[hide]

Skew-symmetric · Skyline · Sparse · Sylvester · Symmetric · Toeplitz · Triangular · Tridiagonal · Vandermonde · Walsh · Z Exchange · Hilbert · Identity · Lehmer · Of ones · Pascal · Pauli · Redheffer · Shift · Zero Constant Companion · Convergent · Defective · Definite · Diagonalizable · Hurwitz · Positive-definite · Stieltjes Conditions on eigenvalues or eigenvectors Congruent · Idempotent or Projection · Invertible · Involutory · Nilpotent · Normal · Orthogonal · Unimodular · Unipotent · Unitary · Totally unimodular · Satisfying conditions on products or inverses Weighing Adjugate · Alternating sign · Augmented · Bézout · Carleman · Cartan · Circulant · Cofactor · Commutation · Confusion · Coxeter · Distance · With specific applications Duplication and elimination · Euclidean distance · Fundamental (linear differential equation) · Generator · Gram · Hessian · Householder · Jacobian · Moment · Payoff · Pick · Random · Rotation · Seifert · Shear · Similarity · Symplectic · Totally positive · Transformation Centering · Correlation · Covariance · Design · Doubly stochastic · Fisher information · Hat · Precision · Stochastic · Transition Used in statistics **Used in graph theory** Adjacency · Biadjacency · Degree · Edmonds · Incidence · Laplacian · Seidel adjacency · Tutte

Mathematics portal · List of matrices · Category:Matrices

**Matrix classes** 

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

Cabibbo-Kobayashi-Maskawa · Density · Fundamental (computer vision) · Fuzzy associative · Gamma · Gell-Mann · Hamiltonian · Irregular · Overlap

Jordan normal form · Linear independence · Matrix exponential · Matrix representation of conic sections · Perfect matrix · Pseudoinverse ·

Metzler · Moore · Nonnegative · Pentadiagonal · Permutation · Persymmetric · Polynomial · Quaternionic · Signature · Skew-Hermitian ·

Categories: Machine learning | Statistical classification

**Related terms** 

Used in science and engineering

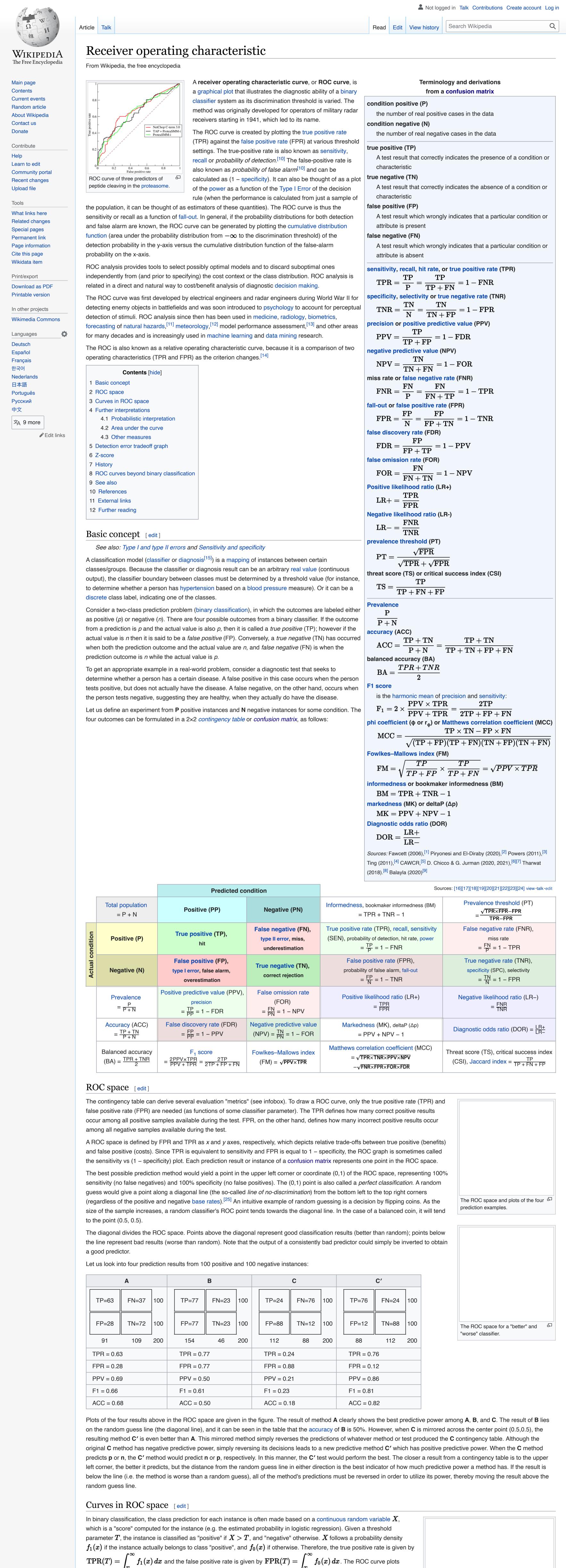
This page was last edited on 31 August 2022, at 12:01 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License 3.0; 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.

· S · State transition · Substitution · Z (chemistry)

Row echelon form · Wronskian

Privacy policy About Wikipedia Disclaimers Contact Wikipedia Mobile view Developers Statistics Cookie statement



• the area between the ROC curve and the no-discrimination line multiplied by two is called the Gini coefficient. It should not be confused with the measure of statistical dispersion also called Gini coefficient. • the area between the full ROC curve and the triangular ROC curve including only (0,0), (1,1) and one selected operating point (tpr, fpr) - Consistency [26] • the area under the ROC curve, or "AUC" ("area under curve"), or A' (pronounced "a-prime"), [27] or "c-statistic" ("concordance statistic"). [28] • the sensitivity index d' (pronounced "d-prime"), the distance between the mean of the distribution of activity in the system under noise-alone conditions and its distribution under signal-alone conditions, divided by their standard deviation, under the assumption that both these distributions are normal with the same standard deviation. Under these assumptions, the shape of the ROC is entirely determined by d'. However, any attempt to summarize the ROC curve into a single number loses information about the pattern of tradeoffs of the particular discriminator algorithm. **Probabilistic interpretation** [edit] When using normalized units, the area under the curve (often referred to as simply the AUC) is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (assuming 'positive' ranks higher than 'negative').[29] In other words, when given one randomly selected positive instance and one randomly selected negative instance, AUC is the probability that the classifier will be able to tell which one is which. This can be seen as follows: the area under the curve is given by (the integral boundaries are reversed as large threshold T has a lower value on the x-axis)

• the intercept of the ROC curve with the tangent at 45 degrees parallel to the no-discrimination line that is closest to the error-free point (0,1) - also called Youden's J statistic

• the intercept of the ROC curve with the line at 45 degrees orthogonal to the no-discrimination line - the balance point where Sensitivity = 1 - Specificity

ᄆ

ᄆ

品

ᄆ

**Mathematics portal** 

35. A Flach, P.A.; Wu, S. (2005). "Repairing concavities in ROC curves." (PDF). 19th

International Joint Conference on Artificial Intelligence (IJCAI'05). pp. 702-707.

areas under receiver operating characteristic curves derived from the same cases" ≥.

36. A Hanley, James A.; McNeil, Barbara J. (1983-09-01). "A method of comparing the

37. A Hanczar, Blaise; Hua, Jianping; Sima, Chao; Weinstein, John; Bittner, Michael;

41. A Hernandez-Orallo, J.; Flach, P.A.; Ferri, C. (2012). "A unified view of performance

42. ^ Powers, David M.W. (2012). "The Problem of Area Under the Curve". International

43. ^ Powers, David M. W. (2003). "Recall and Precision versus the Bookmaker" [m] (PDF).

Proceedings of the International Conference on Cognitive Science (ICSC-2003),

44. ^ Powers, David M. W. (2012). "The Problem with Kappa" (PDF). Conference of the

ROBUS-UNSUP Workshop. Archived from the original (PDF) on 2016-05-18.

45. McClish, Donna Katzman (1989-08-01). "Analyzing a Portion of the ROC Curve".

Medical Decision Making. 9 (3): 190–195. doi:10.1177/0272989X8900900307 ₺.

46. ^ Dodd, Lori E.; Pepe, Margaret S. (2003). "Partial AUC Estimation and Regression" ∠.

Biometrics. 59 (3): 614–623. doi:10.1111/1541-0420.00071 ℃. PMID 14601762 ℃.

University of California, Santa Cruz, in Proceedings of the First International Workshop

47. A Karplus, Kevin (2011); Better than Chance: the importance of null models in,

European Chapter of the Association for Computational Linguistics (EACL2012) Joint

of Machine Learning Research. 13: 2813–2869.

Sydney Australia, 2003, pp. 529–534.

PMID 2668680 ₺. S2CID 24442201 ₺.

Retrieved 2012-07-20.

Retrieved July 11, 2019.

PMID 9673991 2.

PMID 14519861 2.

Conference on Information Science and Technology.

metrics: translating threshold choice into expected classification loss" metrics: translating threshold choice into expected classification loss metrics:

Radiology. 148 (3): 839–843. doi:10.1148/radiology.148.3.6878708

**TOC Curve** 

**ROC Curve** 

Example DET graph

```
A=\int_{-\infty}^{1} 	ext{TPR}(	ext{FPR}^{-1}(x))\,dx=\int_{-\infty}^{-\infty} 	ext{TPR}(T)	ext{FPR}'(T)\,dT=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}I(T'>T)f_1(T')f_0(T)\,dT'\,dT=P(X_1>X_0)
where X_1 is the score for a positive instance and X_0 is the score for a negative instance, and f_0 and f_1 are probability densities as defined in previous section.
```

fusion is much more likely to overfit the data. [35]

Cohen Kappa and Fleiss Kappa. [citation needed][44]

false alarms

false alarms + correct rejections

Detection error tradeoff graph [edit]

dubbed "double probability paper". [52]

Z-score [edit]

have a slope close to 1.0.<sup>[57]</sup>

History [edit]

ROC.<sup>[50]</sup>

Area under the curve [edit]

 $\mathrm{TPR}(T): T \mapsto y(x)$ 

 $\operatorname{FPR}(T):T\mapsto x$ 

 $AUC(f) = rac{\sum_{t_0 \in \mathcal{D}^0} \, \sum_{t_1 \in \mathcal{D}^1} \, \mathbf{1}[f(t_0) < f(t_1)]}{|\mathcal{D}^0| \cdot |\mathcal{D}^1|},$ where,  $\mathbf{1}[f(t_0) < f(t_1)]$  denotes an *indicator function* which returns 1 iff  $f(t_0) < f(t_1)$  otherwise return 0;  $\mathcal{D}^0$  is the set of negative examples, and  $\mathcal{D}^1$  is the set of positive examples.

The AUC is related to the Gini impurity index ( $G_1$ ) by the formula  $G_1 = 2\mathrm{AUC} - 1$ , where:

parametrically  $\mathrm{TPR}(T)$  versus  $\mathrm{FPR}(T)$  with T as the varying parameter.

curve is determined by how much overlap the two distributions have.

Sometimes, the ROC is used to generate a summary statistic. Common versions are:

Further interpretations [edit]

and generalized as Informedness<sup>[citation needed]</sup>

For example, imagine that the blood protein levels in diseased people and healthy people are normally distributed with

classify any number above a certain threshold as indicating disease. The experimenter can adjust the threshold (green

vertical line in the figure), which will in turn change the false positive rate. Increasing the threshold would result in fewer

false positives (and more false negatives), corresponding to a leftward movement on the curve. The actual shape of the

means of 2 g/dL and 1 g/dL respectively. A medical test might measure the level of a certain protein in a blood sample and

 $G_1 = 1 - \sum_{k=1}^n (X_k - X_{k-1})(Y_k + Y_{k-1})^{[33]}$ In this way, it is possible to calculate the AUC by using an average of a number of trapezoidal approximations.  $G_1$  should not be confused with the measure of statistical dispersion that is also called Gini coefficient.

It is also common to calculate the Area Under the ROC Convex Hull (ROC AUCH = ROCH AUC) as any point on the line segment between two prediction results can be

achieved by randomly using one or the other system with probabilities proportional to the relative length of the opposite component of the segment. [34] It is also possible to invert

concavities - just as in the figure the worse solution can be reflected to become a better solution; concavities can be reflected in any line segment, but this more extreme form of

The machine learning community most often uses the ROC AUC statistic for model comparison. [36] This practice has been questioned because AUC estimates are quite noisy

and suffer from other problems. [37][38][39] Nonetheless, the coherence of AUC as a measure of aggregated classification performance has been vindicated, in terms of a uniform

Another problem with ROC AUC is that reducing the ROC Curve to a single number ignores the fact that it is about the tradeoffs between the different systems or performance

performance, and -1 represents the "perverse" case of full informedness always giving the wrong response. [43] Bringing chance performance to 0 allows these alternative scales

to be interpreted as Kappa statistics. Informedness has been shown to have desirable characteristics for Machine Learning versus other common definitions of Kappa such as

Sometimes it can be more useful to look at a specific region of the ROC Curve rather than at the whole curve. It is possible to compute partial AUC. [45] For example, one could

points plotted and not the performance of an individual system, as well as ignoring the possibility of concavity repair, so that related alternative measures such as

It can be shown that the AUC is closely related to the Mann-Whitney U, [30][31] which tests whether positives are ranked higher than negatives. It is also equivalent to the

Wilcoxon test of ranks. [31] For a predictor f, an unbiased estimator of its AUC can be expressed by the following Wilcoxon-Mann-Whitney statistic: [32]

Informedness [citation needed] or DeltaP are recommended. These measures are essentially equivalent to the Gini for a single prediction point with DeltaP' = Informedness = 2AUC-1, whilst DeltaP = Markedness represents the dual (viz. predicting the prediction from the real class) and their geometric mean is the Matthews correlation coefficient. [citation needed] Whereas ROC AUC varies between 0 and 1 — with an uninformative classifier yielding 0.5 — the alternative measures known as Informedness, [citation needed] Certainty [26] and Gini Coefficient (in the single parameterization or single system case)[citation needed] all have the advantage that 0 represents chance performance whilst 1 represents perfect

rate distribution, [40] and AUC has been linked to a number of other performance metrics such as the Brier score. [41]

focus on the region of the curve with low false positive rate, which is often of prime interest for population screening tests. [46] Another common approach for classification problems in which P « N (common in bioinformatics applications) is to use a logarithmic scale for the x-axis. [47] The ROC area under the curve is also called **c-statistic** or **c statistic**.<sup>[48]</sup> Other measures [edit]

The Total Operating Characteristic (TOC) also characterizes diagnostic ability while revealing more information than the ROC. For each

threshold.<sup>[49]</sup> The TOC method reveals all of the information that the ROC method provides, plus additional important information that

ROC does not reveal, i.e. the size of every entry in the contingency table for each threshold. TOC also provides the popular AUC of the

These figures are the TOC and ROC curves using the same data and thresholds. Consider the point that corresponds to a threshold of

An alternative to the ROC curve is the detection error tradeoff (DET) graph, which plots the false negative rate (missed detections) vs.

the false positive rate (false alarms) on non-linearly transformed x- and y-axes. The transformation function is the quantile function of the

that the complement of the hit rate, the miss rate or false negative rate, is used. This alternative spends more graph area on the region of

which, because of using miss rate instead of its complement, the hit rate, is the lower left corner in a DET plot. Furthermore, DET graphs

automatic speaker recognition community, where the name DET was first used. The analysis of the ROC performance in graphs with this

normal distribution, i.e., the inverse of the cumulative normal distribution. It is, in fact, the same transformation as zROC, below, except

interest. Most of the ROC area is of little interest; one primarily cares about the region tight against the y-axis and the top left corner –

have the useful property of linearity and a linear threshold behavior for normal distributions. [51] The DET plot is used extensively in the

warping of the axes was used by psychologists in perception studies halfway through the 20th century, [citation needed] where this was

threshold, ROC reveals two ratios, TP/(TP + FN) and FP/(FP + TN). In other words, ROC reveals  $\frac{---}{\text{hits} + \text{misses}}$ 

74. The TOC curve shows the number of hits, which is 3, and hence the number of misses, which is 7. Additionally, the TOC curve shows that the number of false alarms is 4 and the number of correct rejections is 16. At any given point in the ROC curve, it is possible to glean  $\frac{false~alarms}{false~alarms + correct~rejections}~~\text{and}~~\frac{hits}{hits + misses}.~\text{For example, at threshold 74, it is evident that the x}$ values for the ratios of coordinate is 0.2 and the y coordinate is 0.3. However, these two values are insufficient to construct all entries of the underlying two-bytwo contingency table.

If a standard score is applied to the ROC curve, the curve will be transformed into a straight line. [53] This z-score is based on a normal distribution with a mean of zero and a

standard deviation of one. In memory strength theory, one must assume that the zROC is not only linear, but has a slope of 1.0. The normal distributions of targets (studied

The linearity of the zROC curve depends on the standard deviations of the target and lure strength distributions. If the standard deviations are equal, the slope will be 1.0. If the

has been found that the zROC curve slopes constantly fall below 1, usually between 0.5 and 0.9. [54] Many experiments yielded a zROC slope of 0.8. A slope of 0.8 implies that

have a predicted slope of 1. However, when adding the recollection component, the zROC curve will be concave up, with a decreased slope. This difference in shape and slope

result from an added element of variability due to some items being recollected. Patients with anterograde amnesia are unable to recollect, so their Yonelinas zROC curve would

The ROC curve was first used during World War II for the analysis of radar signals before it was employed in signal detection theory. [58] Following the attack on Pearl Harbor in

1941, the United States army began new research to increase the prediction of correctly detected Japanese aircraft from their radar signals. For these purposes they measured

In the 1950s, ROC curves were employed in psychophysics to assess human (and occasionally non-human animal) detection of weak signals. [58] In medicine, ROC analysis has

been extensively used in the evaluation of diagnostic tests. [60][61] ROC curves are also used extensively in epidemiology and medical research and are frequently mentioned in

conjunction with evidence-based medicine. In radiology, ROC analysis is a common technique to evaluate new radiology techniques. [62] In the social sciences, ROC analysis is

often called the ROC Accuracy Ratio, a common technique for judging the accuracy of default probability models. ROC curves are widely used in laboratory medicine to assess

ROC curves also proved useful for the evaluation of machine learning techniques. The first application of ROC in machine learning was by Spackman who demonstrated the

standard deviation of the target strength distribution is larger than the standard deviation of the lure strength distribution, then the slope will be smaller than 1.0. In most studies, it

On the other hand, TOC shows the total information in the contingency table for each

the variability of the target strength distribution is 25% larger than the variability of the lure strength distribution.<sup>[55]</sup> Another variable used is d'(d prime) (discussed above in "Other measures"), which can easily be expressed in terms of z-values. Although d' is a commonly used parameter, it must be recognized that it is only relevant when strictly adhering to the very strong assumptions of strength theory made above. [56] The z-score of an ROC curve is always linear, as assumed, except in special situations. The Yonelinas familiarity-recollection model is a two-dimensional account of recognition memory. Instead of the subject simply answering yes or no to a specific input, the subject gives the input a feeling of familiarity, which operates like the original ROC curve. What changes, though, is a parameter for Recollection (R). Recollection is assumed to be all-or-none, and it trumps familiarity. If there were no recollection component, zROC would

the ability of a radar receiver operator to make these important distinctions, which was called the Receiver Operating Characteristic. [59]

the diagnostic accuracy of a test, to choose the optimal cut-off of a test and to compare diagnostic accuracy of several tests.

value of ROC curves in comparing and evaluating different classification algorithms. [63]

the sum of the c selected scores over all c! possible ways to assign exactly one example to each class.

1. \* Fawcett, Tom (2006). "An Introduction to ROC Analysis" in (PDF). Pattern

Recognition Letters. 27 (8): 861–874. doi:10.1016/j.patrec.2005.10.010 ℃.

Infrastructure Systems. 26 (1): 04019036. doi:10.1061/(ASCE)IS.1943-

2. A Piryonesi S. Madeh; El-Diraby Tamer E. (2020-03-01). "Data Analytics in Asset

Management: Cost-Effective Prediction of the Pavement Condition Index". Journal of

3. ^ Powers, David M. W. (2011). "Evaluation: From Precision, Recall and F-Measure to

7. ^ Chicco D.; Toetsch N.; Jurman G. (February 2021). "The Matthews correlation

1-22. doi:10.1186/s13040-021-00244-z ₽. PMC 7863449 . PMID 33541410 ₽.

9. A Balayla, Jacques (2020). "Prevalence threshold (φe) and the geometry of screening

8. \* Tharwat A. (August 2018). "Classification assessment methods" . Applied

10. ^ a b "Detector Performance Analysis Using ROC Curves - MATLAB & Simulink

11. A Peres, D. J.; Cancelliere, A. (2014-12-08). "Derivation and evaluation of landslide-

triggering thresholds by a Monte Carlo approach" ∠. Hydrol. Earth Syst. Sci. 18 (12):

4913-4931. Bibcode:2014HESS...18.4913P \(\mathbb{L}\). doi:10.5194/hess-18-4913-2014

12. ^ Murphy, Allan H. (1996-03-01). "The Finley Affair: A Signal Event in the History of

19. ^ Powers, David M. W. (2011). "Evaluation: From Precision, Recall and F-Measure to

ROC, Informedness, Markedness & Correlation" ... Journal of Machine Learning

20. ^ Ting, Kai Ming (2011). Sammut, Claude; Webb, Geoffrey I. (eds.). Encyclopedia of

21. A Brooks, Harold; Brown, Barb; Ebert, Beth; Ferro, Chris; Jolliffe, Ian; Koh, Tieh-Yong;

22. ^ Chicco D, Jurman G (January 2020). "The advantages of the Matthews correlation

BMC Genomics. 21 (1): 6-1–6-13. doi:10.1186/s12864-019-6413-7 △. PMC 6941312

coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness,

and markedness in two-class confusion matrix evaluation" ∠. BioData Mining. 14 (13):

1-22. doi:10.1186/s13040-021-00244-z2. PMC 7863449 . PMID 335414102.

25. ^ "classification - AUC-ROC of a random classifier" ∠. Data Science Stack Exchange.

Consistency and Certainty (PDF). Spring Congress on Engineering and Technology

Research. World Meteorological Organisation. Retrieved 2019-07-17.

23. ^ Chicco D, Toetsch N, Jurman G (February 2021). "The Matthews correlation

24. ^ Tharwat A. (August 2018). "Classification assessment methods" ☑. Applied

26. ^ a b c Powers, David MW (2012). "ROC-ConCert: ROC-Based Measurement of

Computing and Informatics. doi:10.1016/j.aci.2018.08.003

(SCET). Vol. 2. IEEE. pp. 238-241. [dead link]

original (PDF) on 2008-11-20.

External links [edit]

another ROC demo ☑

ROC video explanation ☑

TOC R package on Github ☑

How to run the TOC Package in R ☑

Excel Workbook for generating TOC curves ☑

ROC demo ☑

machine learning. Springer. doi:10.1007/978-0-387-30164-8 ₺. ISBN 978-0-387-30164-

Roebber, Paul; Stephenson, David (2015-01-26). "WWRP/WGNE Joint Working Group

on Forecast Verification Research" . Collaboration for Australian Weather and Climate

13. A Peres, D. J.; Juppa, C.; Cavallaro, L.; Cancelliere, A.; Foti, E. (2015-10-01).

curves" ∠. PLoS One. 15 (10). doi:10.1371/journal.pone.0240215 ∠.

Computing and Informatics. doi:10.1016/j.aci.2018.08.003

Example" ... www.mathworks.com. Retrieved 11 August 2016.

Forecast Verification" . Weather and Forecasting. 11 (1): 3-20.

Bibcode:1996WtFor..11....3M 2. doi:10.1175/1520-

coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness,

and markedness in two-class confusion matrix evaluation" ∠. BioData Mining. 14 (13):

ROC curves are also used in verification of forecasts in meteorology. [64]

ROC curves beyond binary classification [edit]

error variance of the regression model.

Coefficient of determination

• Constant false alarm rate

Detection error tradeoff

Detection theory

References [edit]

555X.0000512℃.

ISSN 1607-7938 ₺.

Technologies. 2 (1): 37–63.

PMID 31898477 ₺.

Retrieved 2020-11-30.

F1 score

See also [edit]

Brier score

objects that the subjects need to recall) and lures (non studied objects that the subjects attempt to recall) is the factor causing the zROC to be linear.

The extension of ROC curves for classification problems with more than two classes is cumbersome. Two common approaches for when there are multiple classes are (1) average over all pairwise AUC values [65] and (2) compute the volume under surface (VUS). [66][67] To average over all pairwise classes, one computes the AUC for each pair of classes, using only the examples from those two classes as if there were no other classes, and then averages these AUC values over all possible pairs. When there are c classes there will be c(c-1)/2 possible pairs of classes. The volume under surface approach has one plot a hypersurface rather than a curve and then measure the hypervolume under that hypersurface. Every possible decision rule that one might use for a classifier for c classes can be described in terms of its true positive rates (TPR<sub>1</sub>, ..., TPR<sub>c</sub>). It is this set of rates that defines a point, and the set of all possible decision rules yields a cloud of points that define the hypersurface. With this definition, the VUS is the probability that the classifier will be able to correctly label all c examples when it is given a set that has one randomly selected example from each class. The implementation of a classifier that knows that its input set consists of one example from each class might first compute a goodness-of-fit score for each of the  $c^2$  possible pairings of an example to a class, and then employ the Hungarian algorithm to maximize

Given the success of ROC curves for the assessment of classification models, the extension of ROC curves for other supervised tasks has also been investigated. Notable

False alarm

ROCCET

Hypothesis tests for accuracy

Precision and recall

Sensitivity and specificity

Total operating characteristic

PMID 6878708 ℃.

proposals for regression problems are the so-called regression error characteristic (REC) Curves [68] and the Regression ROC (RROC) curves. [69] In the latter, RROC curves

become extremely similar to ROC curves for classification, with the notions of asymmetry, dominance and convex hull. Also, the area under RROC curves is proportional to the

ROC, Informedness, Markedness & Correlation" ... Journal of Machine Learning Dougherty, Edward R (2010). "Small-sample precision of ROC-related estimates" ∠. Technologies. 2 (1): 37–63. Bioinformatics. 26 (6): 822–830. doi:10.1093/bioinformatics/btq037 4. ^ Ting, Kai Ming (2011). Sammut, Claude; Webb, Geoffrey I. (eds.). Encyclopedia of PMID 20130029 2. machine learning. Springer. doi:10.1007/978-0-387-30164-8 ₺. ISBN 978-0-387-30164-38. ^ Lobo, Jorge M.; Jiménez-Valverde, Alberto; Real, Raimundo (2008). "AUC: a misleading measure of the performance of predictive distribution models". Global 5. A Brooks, Harold; Brown, Barb; Ebert, Beth; Ferro, Chris; Jolliffe, Ian; Koh, Tieh-Yong; Ecology and Biogeography. 17 (2): 145–151. doi:10.1111/j.1466-8238.2007.00358.x ₺. Roebber, Paul; Stephenson, David (2015-01-26). "WWRP/WGNE Joint Working Group" S2CID 15206363 2. on Forecast Verification Research" . Collaboration for Australian Weather and Climate 39. A Hand, David J (2009). "Measuring classifier performance: A coherent alternative to Research. World Meteorological Organisation. Retrieved 2019-07-17. the area under the ROC curve" ∠. Machine Learning. 77: 103–123. 6. ^ Chicco D.; Jurman G. (January 2020). "The advantages of the Matthews correlation doi:10.1007/s10994-009-5119-5 40. ^ Flach, P.A.; Hernandez-Orallo, J.; Ferri, C. (2011). "A coherent interpretation of AUC BMC Genomics. 21 (1): 6-1-6-13. doi:10.1186/s12864-019-6413-7 ℃. PMC 6941312 as a measure of aggregated classification performance." [m] (PDF). Proceedings of the PMID 31898477 2. 28th International Conference on Machine Learning (ICML-11). pp. 657-664.

```
"Significant wave height record extension by neural networks and reanalysis wind
                                                                                                 on Pattern Recognition in Proteomics, Structural Biology and Bioinformatics (PR PS BB
    data". Ocean Modelling. 94: 128–140. Bibcode:2015OcMod..94..128P ☑.
                                                                                                 2011)
    doi:10.1016/j.ocemod.2015.08.002 2.
                                                                                             48. ^ "C-Statistic: Definition, Examples, Weighting and Significance" ☑. Statistics How To.
14. ^ Swets, John A.; Signal detection theory and ROC analysis in psychology and
                                                                                                 August 28, 2016.
    diagnostics: collected papers ☑, Lawrence Erlbaum Associates, Mahwah, NJ, 1996
                                                                                             49. ^ Pontius, Robert Gilmore; Parmentier, Benoit (2014). "Recommendations for using the
15. A Sushkova, Olga; Morozov, Alexei; Gabova, Alexandra; Karabanov, Alexei;
                                                                                                 Relative Operating Characteristic (ROC)". Landscape Ecology. 29 (3): 367-382.
    Illarioshkin, Sergey (2021). "A Statistical Method for Exploratory Data Analysis Based
                                                                                                 doi:10.1007/s10980-013-9984-8 2. S2CID 15924380 2.
    on 2D and 3D Area under Curve Diagrams: Parkinson's Disease Investigation" 

☑.
                                                                                             50. ^ Pontius, Robert Gilmore; Si, Kangping (2014). "The total operating characteristic to
    Sensors. 21 (14): 4700. doi:10.3390/s21144700 . PMC 8309570
                                                                                                 measure diagnostic ability for multiple thresholds". International Journal of
    PMID 34300440 ₺.
                                                                                                  Geographical Information Science. 28 (3): 570–583.
16. Λ Balayla, Jacques (2020). "Prevalence threshold (φe) and the geometry of screening
                                                                                                 doi:10.1080/13658816.2013.862623 ☑. S2CID 29204880 ☑.
    curves" . PLoS One. 15 (10). doi:10.1371/journal.pone.0240215 .
                                                                                             51. A Navratil, J.; Klusacek, D. (2007-04-01). On Linear DETs. 2007 IEEE International
17. * Fawcett, Tom (2006). "An Introduction to ROC Analysis" in (PDF). Pattern
                                                                                                  Conference on Acoustics, Speech and Signal Processing - ICASSP '07. Vol. 4. pp. IV-
    Recognition Letters. 27 (8): 861–874. doi:10.1016/j.patrec.2005.10.010 ℃.
                                                                                                 229-IV-232. doi:10.1109/ICASSP.2007.367205 . ISBN 978-1-4244-0727-9.
18. ^ Piryonesi S. Madeh; El-Diraby Tamer E. (2020-03-01). "Data Analytics in Asset
                                                                                                 S2CID 18173315 ₺.
    Management: Cost-Effective Prediction of the Pavement Condition Index". Journal of
                                                                                             52. ^ Dev P. Chakraborty (December 14, 2017). "double+probability+paper"&pg=PT214
    Infrastructure Systems. 26 (1): 04019036. doi:10.1061/(ASCE)IS.1943-
                                                                                                  Observer Performance Methods for Diagnostic Imaging: Foundations, Modeling, and
    555X.0000512℃.
```

ROC curve analysis for sensor-based estimates in human computer interaction" \( \mathref{Z} \). ACM International Conference Proceeding Series, Proceedings of Graphics Interface *2005.* Waterloo, ON: Canadian Human-Computer Communications Society. 28. A Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome H. (2009). The elements of statistical learning: data mining, inference, and prediction (2nd ed.). 29. \* Fawcett, Tom (2006); An introduction to ROC analysis , Pattern Recognition Letters, 27, 861-874.

30. A Hanley, James A.; McNeil, Barbara J. (1982). "The Meaning and Use of the Area

doi:10.1148/radiology.143.1.7063747 2. PMID 7063747 2. S2CID 10511727 2.

31. ^ a b Mason, Simon J.; Graham, Nicholas E. (2002). "Areas beneath the relative

Society. 128 (584): 2145–2166. Bibcode:2002QJRMS.128.2145M €.

42–53. doi:10.1007/978-3-540-74976-9\_8 . ISBN 978-3-540-74976-9.

under a Receiver Operating Characteristic (ROC) Curve". Radiology. 143 (1): 29–36.

operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical

significance and interpretation" in (PDF). Quarterly Journal of the Royal Meteorological

CiteSeerX 10.1.1.458.8392 . doi:10.1256/003590002320603584 \angle . Archived from the

32. ^ Calders, Toon; Jaroszewicz, Szymon (2007). Kok, Joost N.; Koronacki, Jacek; Lopez

"Efficient AUC Optimization for Classification" ☑. Knowledge Discovery in Databases:

PKDD 2007. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. 4702:

33. A Hand, David J.; and Till, Robert J. (2001); A simple generalization of the area under

de Mantaras, Ramon; Matwin, Stan; Mladenič, Dunja; Skowron, Andrzej (eds.).

27. ^ Fogarty, James; Baker, Ryan S.; Hudson, Scott E. (2005). "Case studies in the use of

the ROC curve for multiple class classification problems, Machine Learning, 45, 171-186. 34. ^ Provost, F.; Fawcett, T. (2001). "Robust classification for imprecise environments". Machine Learning. **42** (3): 203–231. arXiv:cs/0009007 doi:10.1023/a:1007601015854 2. S2CID 5415722 2.

An Introduction to the Total Operating Characteristic: Utility in Land Change Model Evaluation ☑

• Green, William H., (2003) *Econometric Analysis*, fifth edition, Prentice Hall, ISBN 0-13-066189-9

337–344. doi:10.1111/j.0006-341x.2000.00337.x ₺. PMID 10877287 ₺. S2CID 8822160 ₺.

Transactions in GIS. 7 (4): 467–484. doi:10.1111/1467-9671.00159 ₺. S2CID 14452746 ₺.

Ecosystems & Environment. 85 (1–3): 239–248. doi:10.1016/S0167-8809(01)00187-6 ℃.

Further reading [edit] Balakrishnan, Narayanaswamy (1991); Handbook of the Logistic Distribution, Marcel Dekker, Inc., ISBN 978-0-8247-8587-1 Brown, Christopher D.; Davis, Herbert T. (2006). "Receiver operating characteristic curves and related decision measures: a tutorial". Chemometrics and Intelligent Laboratory Systems. 80: 24–38. doi:10.1016/j.chemolab.2005.05.004 ℃. • Rotello, Caren M.; Heit, Evan; Dubé, Chad (2014). "When more data steer us wrong: replications with the wrong dependent measure perpetuate erroneous conclusions" 📠

CiteSeerX 10.1.1.145.4649 . doi:10.1016/j.patrec.2005.10.012 2.

doi:10.1007/s10980-013-9984-8 2. S2CID 15924380 2.

(4): 325–334. doi:10.1007/s10708-004-5049-5 2. S2CID 155073463 2.

receiver-operating characteristic in recognition memory". Journal of Experimental Psychology: Learning, Memory, and Cognition. 25 (2): 500-513. doi:10.1037/0278-7393.25.2.500 ₺. 55. A Ratcliff, Roger; McCoon, Gail; Tindall, Michael (1994). "Empirical generality of data from recognition memory ROC functions and implications for GMMs". Journal of Experimental Psychology: Learning, Memory, and Cognition. 20 (4): 763–785. CiteSeerX 10.1.1.410.2114 . doi:10.1037/0278-7393.20.4.763 \alpha. 56. A Zhang, Jun; Mueller, Shane T. (2005). "A note on ROC analysis and non-parametric estimate of sensitivity". Psychometrika. 70: 203–212. CiteSeerX 10.1.1.162.1515 doi:10.1007/s11336-003-1119-8 2. S2CID 122355230 2. 57. A Yonelinas, Andrew P.; Kroll, Neal E. A.; Dobbins, Ian G.; Lazzara, Michele; Knight, Robert T. (1998). "Recollection and familiarity deficits in amnesia: Convergence of

Applications with R-Based Examples 

CRC Press. p. 214. ISBN 9781351230711.

53. ^ MacMillan, Neil A.; Creelman, C. Douglas (2005). Detection Theory: A User's Guide

(2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates. ISBN 978-1-4106-1114-7.

remember-know, process dissociation, and receiver operating characteristic data".

psychophysics. New York, NY: John Wiley and Sons Inc. ISBN 978-0-471-32420-1.

59. ^ "Using the Receiver Operating Characteristic (ROC) curve to analyze a classification

60. ^ Zweig, Mark H.; Campbell, Gregory (1993). "Receiver-operating characteristic (ROC)

61. ^ Pepe, Margaret S. (2003). The statistical evaluation of medical tests for classification

62. ^ Obuchowski, Nancy A. (2003). "Receiver operating characteristic curves and their

use in radiology". Radiology. 229 (1): 3–8. doi:10.1148/radiol.2291010898 ℃.

63. ^ Spackman, Kent A. (1989). "Signal detection theory: Valuable tools for evaluating

inductive learning". Proceedings of the Sixth International Workshop on Machine

64. ^ Kharin, Viatcheslav (2003). "On the ROC score of probability forecasts" ☑. Journal of

Climate. 16 (24): 4145–4150. Bibcode:2003JCli...16.4145K 2. doi:10.1175/1520-

65. ^ Till, D.J.; Hand, R.J. (2001). "A Simple Generalisation of the Area Under the ROC

plots: a fundamental evaluation tool in clinical medicine" [a] (PDF). Clinical Chemistry.

University of Utah. Department of Mathematics, University of Utah. Archived in (PDF)

model: A final note of historical interest [ [mail] (PDF). Department of Mathematics,

Neuropsychology. 12 (3): 323-339. doi:10.1037/0894-4105.12.3.323 2.

58. ^ a b Green, David M.; Swets, John A. (1966). Signal detection theory and

**39** (8): 561–577. doi:10.1093/clinchem/39.4.561 . PMID 8472349 ₺.

and prediction. New York, NY: Oxford. ISBN 978-0-19-856582-6.

Learning. San Mateo, CA: Morgan Kaufmann. pp. 160–163.

0442(2003)016<4145:OTRSOP>2.0.CO;2

from the original on 2020-08-22. Retrieved May 25, 2017.

54. A Glanzer, Murray; Kisok, Kim; Hilford, Andy; Adams, John K. (1999). "Slope of the

Curve for Multiple Class Classification Problems" 

∠. Machine Learning. 45 (2): 171– 186. doi:10.1023/A:1010920819831 66. Mossman, D. (1999). "Three-way ROCs". *Medical Decision Making*. **19** (1): 78–89. doi:10.1177/0272989x9901900110 ₽. PMID 9917023 ₽. S2CID 24623127 ₽. 67. ^ Ferri, C.; Hernandez-Orallo, J.; Salido, M.A. (2003). "Volume under the ROC Surface for Multi-class Problems". Machine Learning: ECML 2003. pp. 108-120. 68. A Bi, J.; Bennett, K.P. (2003). "Regression error characteristic curves" (PDF). Twentieth International Conference on Machine Learning (ICML-2003). Washington, DC.

69. A Hernandez-Orallo, J. (2013). "ROC curves for regression". Pattern Recognition. 46

(12): 3395–3411. doi:10.1016/j.patcog.2013.06.014 \(\mathbb{L}\). hdl:10251/40252

- Wikimedia Commons has media related to *Receiver operating* characteristic. (PDF). Psychonomic Bulletin & Review. 22 (4): 944–954. doi: 10.3758/s13423-014-0759-2 . PMID 25384892 ₺. S2CID 6046065 ₺.
- Lasko, Thomas A.; Bhagwat, Jui G.; Zou, Kelly H.; Ohno-Machado, Lucila (2005). "The use of receiver operating characteristic curves in biomedical informatics". Journal of Biomedical Informatics. 38 (5): 404–415. CiteSeerX 10.1.1.97.9674 . doi:10.1016/j.jbi.2005.02.008 ₺. PMID 16198999 ₺. • Mas, Jean-François; Filho, Britaldo Soares; Pontius, Jr, Robert Gilmore; Gutiérrez, Michelle Farfán; Rodrigues, Hermann (2013). "A suite of tools for ROC analysis of spatial models" ☑. ISPRS International Journal of Geo-Information. 2 (3): 869–887. Bibcode:2013IJGI....2..869M ☑. doi:10.3390/ijgi2030869 • Pontius, Jr, Robert Gilmore; Parmentier, Benoit (2014). "Recommendations for using the Relative Operating Characteristic (ROC)" . Landscape Ecology. 29 (3): 367–382.

Clinical Chemistry. 49 (3): 433–439. doi:10.1373/49.3.433 . PMID 12600955 ₺. • Swets, John A.; Dawes, Robyn M.; and Monahan, John (2000); Better Decisions through Science, Scientific American, October, pp. 82–87 • Zou, Kelly H.; O'Malley, A. James; Mauri, Laura (2007). "Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models" . Circulation. 115 (5): 654–7. doi:10.1161/circulationaha.105.594929 . PMID 17283280 ₺. • Zhou, Xiao-Hua; Obuchowski, Nancy A.; McClish, Donna K. (2002). Statistical Methods in Diagnostic Medicine. New York, NY: Wiley & Sons. ISBN 978-0-471-34772-9. V •T •E **Statistics** [hide] Outline · Index **Descriptive statistics** [show] **Data collection** [show] Statistical inference [show] [show] **Correlation · Regression analysis** Categorical / Multivariate / Time-series / Survival analysis **Applications** [show] Category · Mathematics portal · Commons · WikiProject **Machine learning evaluation metrics** [show] V •T •E **Public health** [show] V •T •E **Authority control: National libraries** Germany ∠

• Fawcett, Tom (2004). "ROC Graphs: Notes and Practical Considerations for Researchers" (PDF). Pattern Recognition Letters. 27 (8): 882–891.

• Heagerty, Patrick J.; Lumley, Thomas; Pepe, Margaret S. (2000). "Time-dependent ROC Curves for Censored Survival Data and a Diagnostic Marker". Biometrics. 56 (2):

• Pontius, Jr, Robert Gilmore; Pacheco, Pablo (2004). "Calibration and validation of a model of forest disturbance in the Western Ghats, India 1920–1990" . GeoJournal. 61

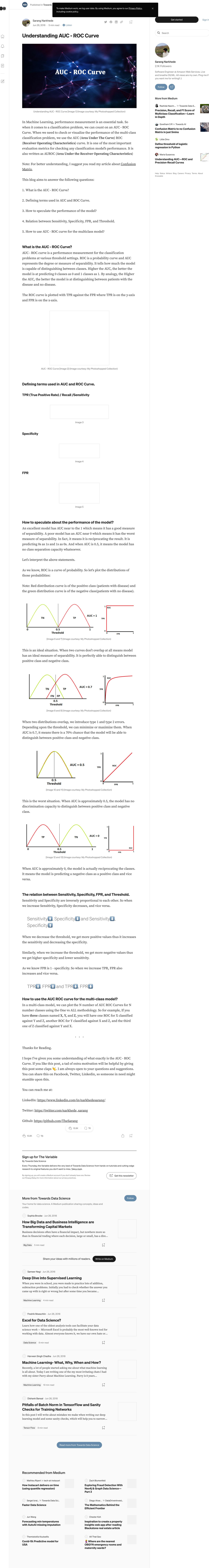
• Pontius, Jr, Robert Gilmore; Batchu, Kiran (2003). "Using the relative operating characteristic to quantify certainty in prediction of location of land cover change in India".

• Pontius, Jr, Robert Gilmore; Schneider, Laura (2001). "Land-use change model validation by a ROC method for the Ipswich watershed, Massachusetts, USA" . Agriculture,

Gonen, Mithat (2007); Analyzing Receiver Operating Characteristic Curves Using SAS, SAS Press, ISBN 978-1-59994-298-8

Hosmer, David W.; and Lemeshow, Stanley (2000); Applied Logistic Regression, 2nd ed., New York, NY: Wiley, ISBN 0-471-35632-8

Categories: Detection theory | Data mining | Biostatistics | Statistical classification | Summary statistics for contingency tables This page was last edited on 22 August 2022, at 18:22 (UTC). Text is available under the Creative Commons Attribution-ShareAlike License 3.0; 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. Privacy policy About Wikipedia Disclaimers Contact Wikipedia Mobile view Developers Statistics Cookie statement



Get started

Sign In