RESEARCH NOTE

Optimal threshold estimation for binary classifiers using game theory [version 1; peer review: 2 approved]

Ignacio Enrique Sanchez

Protein Physiology Laboratory, University of Buenos Aires, Buenos Aires, Argentina

V1 **First published:** 25 Nov 2016, 5(ISCB Comm J):2762
<https://doi.org/10.12688/f1000research.10114.1>
Second version: 15 Dec 2016, 5(ISCB Comm J):2762
<https://doi.org/10.12688/f1000research.10114.2>
Latest published: 08 Feb 2017, 5(ISCB Comm J):2762
<https://doi.org/10.12688/f1000research.10114.3>

Abstract

Many bioinformatics algorithms can be understood as binary classifiers. They are usually trained by maximizing the area under the receiver operating characteristic (ROC) curve. On the other hand, choosing the best threshold for practical use is a complex task, due to uncertain and context-dependent skews in the abundance of positives in nature and in the yields/costs for correct/incorrect classification. We argue that considering a classifier as a player in a zero-sum game allows us to use the minimax principle from game theory to determine the optimal operating point. The proposed classifier threshold corresponds to the intersection between the ROC curve and the descending diagonal in ROC space and yields a minimax accuracy of 1-FPR. Our proposal can be readily implemented in practice, and reveals that the empirical condition for threshold estimation of "specificity equals sensitivity" maximizes robustness against uncertainties in the abundance of positives in nature and classification costs.

Keywords

Binary classifier, ROC curve, accuracy, optimal threshold, optimal cutoff, class imbalance, game theory, minimax principle.

This article is included in the **Bioinformatics** gateway.

This article is included in the **Machine learning: life sciences** collection.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 3			
(revision)		✓	✓
08 Feb 2017			view
version 2			
(revision)			?
15 Dec 2016			view
version 1	✓	✓	
25 Nov 2016	view	view	

1. **Pieter Meysman** , University of Antwerp, Edegem, Belgium
2. **Luis Diambra** , Universidad Nacional de La Plata (UNLP-CONICET), La Plata, Argentina
3. **Marcel Brun**, Universidad Nacional de La Plata (UNLP-CONICET), La Plata, Argentina
Arjen ten Have, Universidad Nacional de La Plata (UNLP-CONICET), La Plata, Argentina

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Ignacio Enrique Sanchez (isanchez@qb.fcen.uba.ar)

Competing interests: No competing interests were disclosed.

Grant information: ANPCyT [PICT 2012-2550]. IES is a CONICET career investigator.

Copyright: © 2016 Sanchez IE. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Sanchez IE. **Optimal threshold estimation for binary classifiers using game theory [version 1; peer review: 2 approved]** F1000Research 2016, 5(ISCB Comm J):2762 <https://doi.org/10.12688/f1000research.10114.1>

First published: 25 Nov 2016, 5(ISCB Comm J):2762 <https://doi.org/10.12688/f1000research.10114.1>

Introduction

Many bioinformatics algorithms can be understood as binary classifiers, as they are used to investigate whether a query entity belongs to a certain class¹. Score-based binary classifiers assign a number to the query. If this score surpasses a threshold, the query is assigned to the class under consideration. A minority of users are able to choose a threshold using their understanding of the algorithm, while the majority uses the default threshold.

Binary classifiers are often trained and compared under a unified framework, the receiver operating characteristic (ROC) curve². Briefly, classifier output is first compared to a training set at all possible classification thresholds, yielding the confusion matrix with the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) (Table 1). The ROC curve plots the true positive rate ($TPR = TP/(TP + FN)$), also called sensitivity, against the false positive rate ($FPR = FP/(FP + TN)$), which equals 1-specificity (Figure 1, continuous line). Classifier training often aims at maximizing the area under the ROC curve, which amounts to maximizing the probability that a randomly chosen positive is ranked before a randomly chosen negative². This summary statistic measures performance without committing to a threshold.

Table 1. Confusion matrix for training of a binary classifier. TP: Number of true positives. FP: Number of false positives. FN: Number of false negatives. TN: Number of true negatives.

		Training set	
		p	n
Classifier output	p'	TP	FP
	n'	FN	TN

Figure 1. Optimal threshold estimation in ROC space for a binary classifier using game theory. The descending diagonal $TPR = 1 - FPR$ (dashed line) minimizes classifier performance with respect to q_p . The intersection between the receiver operating characteristic (ROC) curve (continuous line) and this diagonal maximizes this minimal, worst-case utility and determines the optimal operating point according to the minimax principle (empty circle).

Practical application of a classifier requires using a threshold-dependent performance measure to choose the operating point^{1,3}. This is in practice a complex task because the application domain may be skewed in two ways⁴. First, for many relevant bioinformatics problems the prevalence of positives in nature $q_p = (TP + FN)/(TP + TN + FP + FN)$ does not necessarily match the training set q_p and is hard to estimate^{2,5}. Second, the yields (or costs) for correct and incorrect classification of positives and negatives in the machine learning paradigm (Y_{TP} , Y_{TN} , Y_{FP} , Y_{FN}) may be different from each other and highly context-dependent^{1,3}. Points in the ROC plane with equal performance are connected by iso-yield lines with a slope, the skew ratio, which is the product of the class skew and the yield skew⁴:

$$SKEW\ RATIO = \frac{q_N \cdot (Y_{FP} + Y_{TN})}{q_P \cdot (Y_{TP} + Y_{FN})} \quad (1)$$

The skew ratio expresses the relative importance of negatives and positives, regardless of the source of the skew⁴. Multiple threshold-dependent performance measures have been proposed and discussed in terms of skew sensitivity^{3,4}, but often not justified from first principles.

Theory

Game theory allows us to consider a binary classifier as a zero-sum game between nature and the classifier⁶. In this game, nature is a player that uses a mixed strategy, with probabilities q_p and $q_N = 1 - q_p$ for positives and negatives, respectively. The algorithm is the second player, and each threshold value corresponds to a mixed strategy with probabilities p_p and p_N for positives and negatives. Two of the four outcomes of the game, TP and TN, favor the classifier, while the remaining two, FP and FN, favor nature. The game payoff matrix (Table 2) displays the four possible outcomes and the corresponding classifier utilities a , b , c and d . The Utility of the classifier within the game is:

$$UTILITY = \frac{a \cdot TP + d \cdot TN + b \cdot FP + c \cdot FN}{TP + TN + FP + FN} \quad (2)$$

The payoff matrix for this zero-sum game corresponds directly to the confusion matrix for the classifier, and the game utilities a , b , c , d correspond to the machine learning yields Y_{TP} , Y_{FP} , Y_{FN} , Y_{TN} respectively (Table 1). Without loss of generality⁴, we can study

Table 2. Payoff matrix for a zero-sum game between nature and a binary classifier.

a : Player I utility for a true positive. b : Player I utility for a false positive. c : Player I utility for a false negative. d : Player I utility for a true negative.

		Player II: Nature	
		p	n
Player I: Classifier	p'	a	b
	n'	c	d

the case $a=d=1$ and $b=c=0$. Classifier *Utility* within the game then reduces to the *Accuracy* or fraction of correct predictions²⁻⁴. In sum, maximizing the *Utility* of a binary classifier in a zero-sum game against nature is equivalent to maximizing its *Accuracy*, a common threshold-dependent performance measure.

We can now use the *minimax* principle from game theory⁶ to choose the operating point for the classifier. This principle maximizes utility for a player within a game using a pessimistic approach. For each possible action a player can take, we calculate a worst-case utility by assuming that the other player will take the action that gives them the highest utility (and the player of interest the lowest). The player of interest should take the action that maximizes this minimal, worst-case utility. Thus, the *minimax* utility of a player is the largest value that the player can be sure to get regardless of the actions of the other player.

In our classifier *versus* nature game, *Utility/Accuracy* of the classifier is skew-sensitive, depending on q_p for a given threshold^{3,4}:

$$UTILITY = 1 - FPR + q_p \cdot (FPR + TPR - 1) \quad (3)$$

The derivative of the *Utility* with respect to q_p is zero along the $TPR = 1 - FPR$ line in ROC space (Figure 1, dashed line). The derivative is negative below this line and positive above it, indicating that points along this line are minima of the *Utility* function with respect to the strategy q_p of the nature player. According to the *minimax* principle, the classifier player should operate at the point along the $TPR = 1 - FPR$ line that maximizes *Utility*. In ROC space, this condition corresponds to the intersection between the ROC curve and the descending diagonal (Figure 1, empty circle) and yields a *minimax* value of $1 - FPR$ for the *Utility*. It is worth noting that this analysis regarding class skew is also valid for yield/cost skew⁴.

Discussion

We showed that binary classifiers may be analyzed in terms of game theory. From the *minimax* principle, we propose a criterion to choose an operating point for the classifier that maximizes robustness against uncertainties in the skew ratio, i.e., in the prevalence of positives in nature and in yield skew, i.e., the yields/costs for true positives, true negatives, false positives and false negatives. This can be of practical value, since these uncertainties are widespread in bioinformatics and clinical applications.

In machine learning theory, $TPR = 1 - FPR$ is the line of skew-indifference for *Accuracy* as a performance metric⁴. This is in agreement with the skew-indifference condition imposed by the *minimax* principle from game theory. However, to our knowledge, skew-indifference has not been exploited for optimal threshold estimation. Furthermore, the operating point of a classifier is often chosen by balancing sensitivity and specificity, without reference to the rationale behind⁷. Our game theory analysis shows that this empirical practice can be understood as a maximization of classifier robustness.

Competing interests

No competing interests were disclosed.

Grant information

ANPCyT [PICT 2012-2550]. IES is a CONICET career investigator.

Acknowledgements

I would like to thank Juan Pablo Pinasco and Francisco Melo for discussion.

References

- Swets JA, Dawes RM, Monahan J: **Better decisions through science**. *Sci Am*. 2000; **283**(4): 82–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Fawcett T: **An introduction to ROC analysis**. *Pattern Recognit Lett*. 2006; **27**(8): 861–874.
[Publisher Full Text](#)
- Okeh UM, Okoro CN: **Evaluating Measures of Indicators of Diagnostic Test Performance: Fundamental Meanings and Formulas**. *J Biomet Biostat*. 2012; **3**(1): 132.
[Publisher Full Text](#)
- Flach PA: **The geometry of ROC space: understanding machine learning metrics through ROC isometrics**. *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*. 2003; 194–201.
[Reference Source](#)
- Tomba M, Li N, Bailey TL, et al.: **Assessing computational tools for the discovery of transcription factor binding sites**. *Nat Biotechnol*. 2005; **23**(1): 137–44.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Von Neumann J, Morgenstern O: **Theory of games and economic behavior**. 6th ed., USA: Princeton university press. 1955.
[Reference Source](#)
- Carmona SJ, Nielsen M, Schafer-Nielsen C, et al.: **Towards High-throughput Immunomics for Infectious Diseases: Use of Next-generation Peptide Microarrays for Rapid Discovery and Mapping of Antigenic Determinants**. *Mol Cell Proteomics*. 2015; **14**(7): 1871–84.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 07 December 2016

<https://doi.org/10.5256/f1000research.10895.r17994>

© 2016 Diambra L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Luis Diambra

Centro Regional de Estudios Genómicos, Universidad Nacional de La Plata (UNLP-CONICET), La Plata, Argentina

The author presents a criterion to choose the operating point for a binary classifier. This criterion is analyzed in term of the game theory. By using the mininax principle author proposes to use as classifier threshold the intersection between the ROC curve and the descending diagonal in ROC space. This operating point for the classifier could maximizes the robustness against some bias in the training set. I found some novelty in the fact to consider such bias for an optimal threshold estimation. The paper is well written and organized but I think also that it could be improved by incorporating some general considerations that helping readers to a better understanding of the problem and the present proposition [1,2].

In the binary classification problem one is trying to deduce the answers to new questions, rather than just recall the answers to old ones. In order to do that we need to train the classifier from question-answer pairs (the training set). This is called supervised learning, because it requires a teacher, knowing the rule, which gives the correct answer to the example questions. In the case here, the author consider score-based binary classifiers, which does not need such learning stage. Could the author put the problem in the context supervised vs. no-supervised?

In the supervised learning context the classifier threshold is a parameter that is found during the learning stage. Training the classifier maximizing the area under ROC curve is an strategy for the classifier learn the training set. Consequently, the proposed strategy could be considered as a "learning rule". However, the performance over new examples is not guaranteed. Other point which can improve the manuscript would be to consider the ability of generalization of the proposed strategy. Could the author add a discussion in this sense?

I believe that this manuscript is of an acceptable scientific standard, and that it will be of interest to a wide audience; however, the manuscript could be revised, as outlined above.

References

1. Watkin T, Rau A, Biehl M: The statistical mechanics of learning a rule. *Reviews of Modern Physics*. 1993; **65** (2): 499-556 [Publisher Full Text](#)
2. Diambra L, Fernández J: Information theory approach to learning of the perceptron rule. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2001; **64** (4 Pt 2): 046106 [PubMed Abstract](#) | [Publisher Full Text](#)

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 06 December 2016

<https://doi.org/10.5256/f1000research.10895.r17996>

© 2016 Meysman P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pieter Meysman

Department of Mathematics and Computer Science, University of Antwerp, Edegem, Belgium

The article by Ignacio Enrique Sanchez concerns a common problem in machine learning, namely the selection of the optimal classification threshold, and provides a mathematical solution based on the principles of game theory. The main concern of the article deals with the unknown distribution of positive and negative samples in the 'real world' or 'nature', thus beyond the provided training data set. The provided derivation is very elegant, and luckily for those researchers in the field the solutions turns out to be to select a threshold where sensitivity and specificity are equal in the training data set.

The biggest concern from the perspective of game theory is that 'nature' is not a conscious agent, and thus will not mischievously choose a positive/negative fraction where the classifier will perform the worst. However as stated in the article, this is to simulate the worst case scenario. However this also means that the threshold calculation may only be optimal in this worst case scenario, but suboptimal in all other cases. It is therefore still not the final word in threshold optimisation, and still leaves machine learning researchers the flexibility to choose other thresholds.

However I do have a minor comment on the derivation, that I expect can be addressed with small clarifications to the text:

1. The *Accuracy* equals the *Utility* as defined by the payoff matrix in the specific case $a=d=1$ and $b=c=0$, which is stated without a loss in generality. However in my understanding, this step makes the assumption that the cost for a false negative and the cost for a false positive is equal, which may not be the case for all classifiers. Thus it is unclear if this specific case can be transposed to all classifiers in general.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Comments on this article

Version 1

Reader Comment 12 Dec 2016

Matúš Kalaš

Just a note, perhaps an interesting complement is the article [Saito & Rehmsmeier 2015](#) and website [Classifier evaluation with imbalanced datasets](#).

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research