Optimal Data Analysis
Vol. 3 (November 14, 2014), 117-120

2155-0182/10/$3.00

UniODA *vs*. ROC Analysis: Computing the "Optimal" Cut-Point

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

*R*eceiver *o*perator *c*haracteristic (ROC) analysis[1] is sometimes used to assess the classification accuracy achieved using an ordered attribute to discriminate a dichotomous class variable, and in this context to identify an "optimal" discriminant cutpoint. In ROC analysis the optimal cutpoint corresponds to the threshold value at which distance from the ROC curve to the point representing perfect classification accuracy is minimized. This note discusses the difference between an ROC-defined optimal discriminant threshold, and the optimal cutpoint identified by UniODA that maximizes ESS for the sample.[2] ROC and UniODA methods are illustrated and compared for an application involving prediction of Cesarean delivery.

The *ROC curve* is a plot that displays sensitivity ("true positive rate") on the ordinate, and 1-specificity ("false positive rate") on the abscissa, for all possible threshold values (cut-points) that separate class 0 and class 1 observa-tions in the sample. The total *a*rea *u*nder the ROC *c*urve (AUC) is used as an index of the classification performance (predictive validity) of scores on the attribute. The greater the AUC value, the greater the ability of scores on the attribute to correctly classify the two class categories for the sample: the AUC is sample-specific—that is, it isn't normed against chance. In ROC analysis the distance d between the point representing perfect classification and any point on the ROC curve is: $d=\sqrt{[(1-s_n)^2+(1-s_p)^2]}$, where s_n=sensitivity and s_p=specificity. The optimal cutpoint for discriminating the class categories in ROC analysis is defined as the threshold value associated with the minimum value of d. In contrast, in UniODA the optimal threshold is defined as the cutpoint maximizing the value $(s_n+s_p)/2$, that yields the maximum possible ESS for the sample: ESS=0 indicates accuracy expected by chance, and ESS=100 indicates perfect discrimination.[2] Because $\sqrt{[(1-s_n)^2+(1-s_p)^2]}$ and $(s_n+s_p)/2$ aren't isomorphic, ROC analysis and UniODA mustn't identify identical optimal discriminant threshold values for a given sample.

These competing methods are illustrated for an application predicting Cesarean delivery (the dichotomous class variable) on the basis of duration of membrane rupture (in hours) for a sample of n=166 hospitalized women.[3] For this sample Table 1 presents every possible cutpoint value (ranging from 0 to 21.5), as well as the corresponding values of sensitivity, 1-specific-ity, specificity, ROC distance measure d, and the UniODA performance measure ESS.

2155-0182/10/$3.00

Table 1: Computing the "Optimal" Cut-Point Value by ROC Analysis versus by UniODA

Optimal Value	Point	Cutpoint	Sensitivity	1-Specificity	Specificity	Distance	ESS
	1	0.00	1.000	1.0000	0.0000	1.00000	0
	2	0.63	1.000	0.9760	0.0240	0.97600	2.40
	3	0.88	1.000	0.9690	0.0310	0.96900	3.10
	4	1.25	1.000	0.8900	0.1100	0.89000	11.00
	5	1.75	1.000	0.8660	0.1340	0.86600	13.40
	6	2.13	1.000	0.8190	0.1810	0.81900	18.10
	7	2.38	1.000	0.8110	0.1890	0.81100	18.90
	8	2.75	1.000	0.7800	0.2200	0.78000	22.00
	9	3.25	1.000	0.7170	0.2830	0.71700	28.30
	10	3.75	1.000	0.7090	0.2910	0.70900	29.10
	11	4.50	1.000	0.6460	0.3540	0.64600	35.40
	12	5.13	0.971	0.5830	0.4170	0.58372	38.80
	13	5.38	0.971	0.5750	0.4250	0.57573	39.60
	14	5.75	0.971	0.5510	0.4490	0.55176	42.00
	15	6.25	0.914	0.3780	0.6220	0.38766	53.60
	16	6.75	0.914	0.3460	0.6540	0.35653	56.80
	17	7.13	0.857	0.2910	0.7090	0.32424	56.60
	18	7.38	0.857	0.2830	0.7170	0.31708	57.40
	19	7.75	0.857	0.2760	0.7240	0.31085	58.10
	20	8.25	0.800	0.1892	0.8108	0.27531	61.08
ROC	21	**8.75**	0.800	0.1810	0.8190	**0.26974**	61.90
	22	9.25	0.743	0.1100	0.8900	0.27955	63.30
ODA	23	**9.75**	0.743	0.1020	0.8980	0.27650	**64.10**
	24	10.25	0.543	0.0390	0.9610	0.45866	50.40
	25	10.75	0.543	0.0310	0.9690	0.45805	51.20
	26	11.50	0.457	0.0240	0.9760	0.54353	43.30
	27	12.50	0.400	0.0080	0.9920	0.60005	39.20
	28	13.50	0.343	0.0000	1.0000	0.65700	34.30
	29	14.50	0.286	0.0000	1.0000	0.71400	28.60
	30	15.50	0.257	0.0000	1.0000	0.74300	25.70
	31	16.50	0.200	0.0000	1.0000	0.80000	20.00
	32	17.50	0.171	0.0000	1.0000	0.82900	17.10
	33	18.50	0.143	0.0000	1.0000	0.85700	14.30
	34	19.50	0.114	0.0000	1.0000	0.88600	11.40
	35	20.25	0.057	0.0000	1.0000	0.94300	5.70
	36	21.50	0.000	0.0000	1.0000	1.00000	0

As seen, the minimum distance d is 0.27 (shown in red), corresponding to an optimal cutpoint of 8.75 for ROC analysis. The optimal ESS value is 64.1 (shown in green), corresponding to an optimal cutpoint of 9.75 for UniODA. ESS is 61.9 for the ROC cutpoint, yielding 3.4% lower classification performance than was obtained using UniODA. These results clearly demonstrate that ESS achieved using this ROC analysis approach is *not* explicitly optimal. Understanding the extent to which maximum ESS is underachieved in the literature requires additional research examining many different applications. However, unless the UniODA approach is employed in any given application, it is clear that the magnitude of suboptimality will remain unknown.

2155-0182/10/$3.00

It should be noted that differential costs of both types of *misclassifications* are important in some applications—such as the diagnosis of disease, for example. Indeed, the same point may also be raised regarding the differential effect of both types of *correct* classifications. UniODA provides the most powerful methodology for including these additional types of considerations.[2]

In the maximum-accuracy paradigm every application and hypothesis (not only ROC analysis) can be weighted using any quantitative index (e.g., desirability, valence, threat, fear, cost, return, price, distance, time, mass, etc.). Differential weights may be assigned to differential class categories: for example, class 0 (negative) observations can be weighted using increasingly large weights (e.g., successive integers) until the desired level of specificity is attained. The same procedure may be used to over-weight class 1 (positive) observations and thereby maximize sensitivity. And, of course, a series of numerical weights for each observation can be multiplied, in order to model the overall interactive effect of the profile of weights.

If an application involves multiple weights for multiple aspects of the decision-making outcomes, then depending on the application a MCDM approach conducted separately by individual, or for a group, might (i.e., possibly could meaningfully) be used to define a gestalt weight, or solutions involving separate weighting strategies can be developed and compared in the context of the application.

Weights can be obtained for and assigned *to the class categories* (UniODA software can analyze problems involving up to ten separate class categories). Alternatively, weights can be applied *individually to each subject* in the study—since not all people have the same weighting priorities.

A tremendous unique advantage of maximum-accuracy methods is that (weighted) classification accuracy is summarized using the normed ESS index on which 0 is the (weighted) accuracy expected by chance, and 100 is perfect (weighted) classification accuracy. In this manner solutions can be meaningfully compared between different applications, by removing the accuracy expected simply by chance from the solutions. It is important to understand that unlike the normed ESS statistic, the maximum overall accuracy solution, (True Positives + True Negatives)/N, is *not* normed against chance.[2]

Probably the most important unique advantage, however, is that unlike all other methods, only novometric maximum-accuracy methods can identify the entire descendant family of possible strata (cutpoints) for an application, rather than being limited to one cutpoint.[4] This is important in applications where there are multiple groups of people who have different weighting strategies. Combining disparate groups and developing only one stratification criterion can result in paradoxical confounding: in the worst cases of confounding the combined solution can represent a very poor solution for all of the different sample strata.[5] Therefore, use of optimal methods ensures the discovery of models that yield maximum (weighted) accuracy for every hypothesis and sample, and the identification of valid models yielding more reproducible findings than models obtained using alternative methods.[6]

References

[1]Hanley JA, McNeil BJ (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29-36.

[2]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*, Washington, DC, APA Books.

[3]http://www.medicalbiostatistics.com/roccurve.pdf

2155-0182/10/$3.00

[4]Yarnold PR, Soltysik RC (2014). Globally optimal statistical classification models, I: Binary class variable, one ordered attribute. *Optimal Data Analysis*, *3*, 55-77.

[5]Yarnold PR (1996). Characterizing and circumventing Simpson's paradox for ordered bivariate data. *Educational and Psychological Measurement*, *56*, 430-442.

[6]Yarnold PR (2014). Increasing the validity and reproducibility of scientific findings. *Optimal Data Analysis*, *3*, 107-109.

Author Notes

E-mail: Journal@OptimalDataAnalysis.com.

Mail: Optimal Data Analysis, LLC
6348 N. Milwaukee Ave., #163
Chicago, IL 60646