

Optimal thresholds criteria for ROC surfaces

C. S. Hong¹ · E. S. Jung²

^{1,2}Department of Statistics, Sungkyunkwan University

Received 20 August 2013, revised 23 September 2013, accepted 1 October 2013

Abstract

Consider the ROC surface which is a generalization of the ROC curve for three-class diagnostic problems. In this work, we propose five criteria for the three-class ROC surface by extending the Youden index, the sum of sensitivity and specificity, the maximum vertical distance, the amended closest-to-(0,1) and the true rate. It may be concluded that these five criteria can be expressed as a function of two Kolmogorov–Smirnov statistics. A paired optimal thresholds could be obtained simultaneously from the ROC surface. It is found that the paired optimal thresholds selected from the ROC surface are equivalent to the two optimal thresholds found from the two ROC curves.

Keywords: Accuracy, classification, cut-off, marker, sensitivity, threshold, true rate.

1. Introduction

Roc curves are commonly used for the evaluation of diagnostic markers in two-class diagnostic testing. For discrimination tasks with more than two events, it is necessary that the proper analysis of multiple-class diagnostic testing requires a generalization of ROC analysis. Recently, ROC analysis has been extended to three-class diagnostic problems (Scurfield, 1996; Mossman, 1999; Dreiseitl *et al.*, 2000; Heckerling, 2001; Fawcett, 2003; Nakas *et al.*, 2004, 2010; Patel and Markey, 2005; Wandishin and Mullen, 2009; and many others).

Consider a discrimination task involving three events, denoted $\{e_1, e_2, e_3\}$ as assigned by the testing procedure, in which an observer attempts to discriminate among a set of three decisions, denoted $\{d_1, d_2, d_3\}$. Let $\{X_{11}, X_{12}, \dots, X_{1n_1}\}$ be a random sample of size n_1 obtained from the first class with the cumulative distribution function $F_1(\cdot)$, and let $\{X_{21}, X_{22}, \dots, X_{2n_2}\}$ and $\{X_{31}, X_{32}, \dots, X_{3n_3}\}$ be random samples of sizes n_2 and n_3 obtained from the second and third classes, $F_2(\cdot)$ and $F_3(\cdot)$, respectively. Assume that $F_1(x) \geq F_2(x) \geq F_3(x)$ for all x . For two ordered thresholds (decision markers, cut-off point) $c_1 \leq c_2$, the following decision rule may be applied:

If $X \leq c_1$ then decision is class 1 (d_1).

Else if $c_1 \leq X < c_2$ then decision is class 2 (d_2).

Else decision is class 3 (d_3).

¹ Corresponding author: Professor, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea. E-mail: cshong@skku.edu

² Graduate student, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea.

Table 1.1 shows the event-decision matrix of the observer. Each entry in the matrix specifies the probability that the observer will make a particular decision given that a particular event occurred. For example, the three probabilities, $P(d_k|e_k)$ ($k = 1, 2, 3$) are the true classification rates, and the six probabilities, $P(d_i|e_j)$, ($i \neq j$) are the false classification rates. Note that

Table 1.1 Event-decision probability matrix

		Event		
		e_1	e_2	e_3
Decision	d_1	$P(d_1 e_1)$	$P(d_1 e_2)$	$P(d_1 e_3)$
	d_2	$P(d_2 e_1)$	$P(d_2 e_2)$	$P(d_2 e_3)$
	d_3	$P(d_3 e_1)$	$P(d_3 e_2)$	$P(d_3 e_3)$

Table 1.2 Event-decision distribution function matrix

		Event		
		e_1	e_2	e_3
Decision	d_1	$F_1(c_1)$	$F_2(c_1)$	$F_3(c_1)$
	d_2	$F_1(c_2) - F_1(c_1)$	$F_2(c_2) - F_2(c_1)$	$F_3(c_2) - F_3(c_1)$
	d_3	$1 - F_1(c_2)$	$1 - F_2(c_2)$	$1 - F_3(c_2)$

$$\begin{aligned}
 P(d_1|e_1) &= P(X_1 < c_1) = F_1(c_1), \\
 P(d_2|e_2) &= P(c_1 \leq X_2 < c_2) = F_2(c_2) - F_2(c_1), \\
 P(d_3|e_3) &= P(c_2 \leq X_3) = 1 - F_3(c_2).
 \end{aligned}$$

Table 1.2 shows that the probabilities, $P(d_i|e_j)$, in Table 1.1 are represented by three cumulative distribution functions $F_j(\cdot)$ and two thresholds $c_1 \leq c_2$.

The surface generated by the three kinds of the true classification rates as the two thresholds vary over the domain of X is called the ROC surface. The functional form of the ROC surface is

$$ROC_s(p_1, p_3) = F_2(F_3^{-1}(1 - p_3)) - F_2(F_1^{-1}(p_1)), \quad 0 \leq p_1, p_3 \leq 1,$$

where $p_1 = P(d_1|e_1)$ and $p_3 = P(d_3|e_3)$. The ROC surface is constructed by plotting the points $(F_1(c_1), F_2(c_2) - F_2(c_1), 1 - F_3(c_2))$ in three dimensional unit-length cube (Nakas and Yiannoutsos., 2004).

Threshold selection in a three-class ROC surface has been explained in a theoretical context in He and Frey (2006). Nakas *et al.* (2010, 2013) generalized the Youden index for the three-class classification problem. In this paper, we extend the generalization of the Youden index and propose several optimal threshold criteria for a three-dimensional ROC surface.

There are four optimal threshold criteria which are related to the Youden index for the ROC curve: the sum of the sensitivity and specificity, the maximum vertical distance, the amended closest-to-(0,1), and the true rate. These four criteria defined on the ROC curve are extended and developed for the ROC surface in Section 2. In Section 3, some examples are illustrated to describe the relation between the paired thresholds obtained from the ROC surface and the two thresholds obtained separately from the two ROC curves. We derive a conclusion in Section 4.

2. Optimal criteria for the ROC surface

There are many criteria used to select optimal thresholds (Youden, 1950; Connell and Koepsell, 1985; Perkins and Schisterman, 2006; Kraznowski and Hand, 2009; Velez *et al.*, 2007; Hong, 2009; and many others). Among them, the well-known Youden index J is defined as the maximum of the simple average of the differences between the true classification rate and the false classification rate for each event, that is, the average of $F_1(x) - (1 - F_1(x))$ and $(1 - F_2(x)) - F_2(x)$. Hence the Youden index is represented as $J = \max[F_1(x) - F_2(x)]$ (Youden, 1950), which is essentially the Kolmogorov–Smirnov statistic.

Consider four optimal criteria for the ROC curve which have a relationship with the Youden index: the sum of the sensitivity and specificity is defined to maximize the summation of the sensitivity and the specificity, which is represented as $SSS = \max[F_1(x) + (1 - F_2(x))]$ (Connell and Koepsell, 1985). The MVD (maximum vertical distance) is defined as the maximum vertical difference between the ROC curve and the chance line: $MVD = \max |ROC(p) - p| = \max[F_1(x) - F_2(x)]$, where $ROC(p) = F_1(F_2^{-1}(p))$, $p \in [0, 1]$ (Kraznowski and Hand, 2009). The amended closest-to-(0,1) criteria, AC , is defined to minimize the proportion of the smaller radius (r_2) (the distance between point (1, 0) and a point on the ROC curve) to the larger radius (r_1) (the distance between point (0, 1) and the chance line), that is, $AC = \min[1 - F_1(x) + F_2(x)]$ (Perkins and Schisterman, 2006). And the balanced accuracy (BA) or the true rate (TR) is defined as the maximum of the arithmetic mean of the sensitivity (true positive rate) and the specificity (true negative rate): $TR = \max [F_1(x) + (1 - F_2(x))]/2$ (Velez *et al.*, 2007; Hong, 2009). Hong and Joo (2010), Hong *et al.* (2010) and Yoo and Hong (2011) found that these five criteria are represented as a function of the Kolmogorov–Smirnov statistic. Hence we make use of these statistics for selecting a paired optimal thresholds based on the ROC surface.

First, Nakas *et al.* (2013) generalized the Youden index to multiple-class classification problems. However, to keep the definition of the Youden index for the ROC curve and extend its use to the three– class ROC surface, we will define a Youden index (J_3) for the ROC surface which maximize the average of the three differences between the one true classification rate and the two false classification rates. And explain that the Youden index for the ROC surface has a relationship with Youden indices and KS statistics for the two ROC curves.

Proposition 2.1 Youden index for ROC surface

$$\begin{aligned} J_3 &= \max_{x,y;x < y} [\{F_1(x) - (F_1(y) - F_1(x)) - (1 - F_1(y))\} + \{(F_2(y) - F_2(x)) \\ &\quad - (F_2(x)) - (1 - F_2(y))\} + \{(1 - F_3(y)) - F_3(x) - (F_3(y) - F_3(x))\}]/3 \\ &= \max_{x,y;x < y} [2 \times F_1(x) + 2 \times (F_2(y) - F_2(x)) + 2 \times (1 - F_3(y))]/3. \end{aligned}$$

Corollary 2.1 Relation of J and KS

$$\begin{aligned} J_3 &= 2/3 \times (\max[J_{12}] + \max[J_{23}]) - 1/3 \\ &= 2/3 \times (KS_{12} + KS_{23}) - 1/3, \end{aligned}$$

where J_{ij} is the Youden index for the two dimensional ROC curve consisting of $F_i(\cdot)$ and $F_j(\cdot)$, and KS_{ij} is the Kolmogorov–Smirnov statistic for testing $H_0 : F_i(\cdot) = F_j(\cdot)$. That is,

$J_{12} = \max[F_1(x) - F_2(x)]$ for all x . And there exists constant, c_1 , which satisfies $F_1(c_1) - F_2(c_1) = \max[F_1(x) - F_2(x)]$ for all x , denoted as KS_{12} .

By extending the sum of the sensitivity and specificity (SSS) to the three-dimensional ROC surface, we define STR_3 to maximize the sum of the three true classification rates and its relationship as follows:

Proposition 2.2 Sum of three true rates for ROC surface

$$STR_3 = \max_{x,y;x < y} [F_1(x) + (F_2(y) - F_2(x)) + (1 - F_3(y))].$$

Corollary 2.2 Relation of SSS and KS

$$\begin{aligned} STR_3 &= SSS_{12} + SSS_{23} - 1 \\ &= KS_{23} + KS_{12} + 1, \end{aligned}$$

where SSS_{ij} is the SSS for the ROC curve consisting of $F_i(\cdot)$ and $F_j(\cdot)$.

The MVD (maximum vertical distance) for the ROC curve can be extended to a three-class ROC surface as the maximum vertical difference between the ROC surface, $ROC_3(p_1, p_3)$, and the chance plane, $g(p_1, p_3) = 1 - p_1 - p_3$.

Proposition 2.3 MVD for ROC surface

$$\begin{aligned} MVD_3 &= \max_{x,y;x < y} [F_2(F_3^{-1}(1 - p_3)) - F_2(F_1^{-1}(p_1)) - (1 - p_1 - p_3)] \\ &= \max_{x,y;x < y} [F_2(y) - F_3(y) + F_1(x) - F_2(x)]. \end{aligned}$$

Corollary 2.3 Relation of MVD and KS

$$\begin{aligned} MVD_3 &= MVD_{12} + MVD_{23} \\ &= KS_{23} + KS_{12}, \end{aligned}$$

where MVD_{ij} is the MVD for ROC curve consisting of $F_i(\cdot)$ and $F_j(\cdot)$.

The amended closest-to-(0,1) criteria can be extended to the amended closest-to-(1,1,1) criteria for the ROC Surface as $\min[\sqrt{r_2^2/r_1^2}]$, where r_1^2 and r_2^2 are obtained as

$$\begin{aligned} r_1 &= \frac{\sqrt{\{(1 - F_1(x))^2 + (1 - (F_2(y) - F_2(x)))^2 + (F_3(y))^2\}}}{1 - (F_1(x) - F_2(x) + F_2(y) - F_3(y))}, \\ r_2 &= \sqrt{\{(1 - F_1(x))^2 + (1 - (F_2(y) - F_2(x)))^2 + (F_3(y))^2\}}. \end{aligned}$$

Hence the amended closest-to-(1,1,1) criteria, AC_3 , can be defined as

Proposition 2.4 Amended closest-to-(1,1,1) criteria for ROC surface

$$\begin{aligned} AC_3 &= \min[\sqrt{r_2^2/r_1^2}] \\ &= \min[1 - (F_1(x) - F_2(x) + F_2(y) - F_3(y))]. \end{aligned}$$

Corollary 2.4 Relation of AC and KS

$$\begin{aligned} AC_3 &= AC_{12} + AC_{23} - 1 \\ &= 1 - (KS_{12} + KS_{23}), \end{aligned}$$

where AC_{ij} is the amended closest-to-(0,1) for the ROC curve consisting of $F_i(\cdot)$ and $F_j(\cdot)$.

The balanced accuracy (BA) or the true rate (TR) can be generalized to the ROC surface as the maximum of the simple mean of the three kinds of true classification rates:

Proposition 2.5 TR for ROC surface

$$TR_3 = \max_{x,y;x < y} [1/3 \times [F_1(x) + (F_2(y) - F_2(x)) + (1 - F_3(y))]].$$

Corollary 2.5 Relation of TR and KS

$$\begin{aligned} TR_3 &= 2/3(TR_{23} + TR_{12} + 1/2) \\ &= 1/3(KS_{12} + KS_{23} + 1), \end{aligned}$$

where the TR_{ij} is the true rate for the ROC curve consisting of $F_i(\cdot)$ and $F_j(\cdot)$.

Therefore the five classification criteria such as J_3 , STR_3 , MVD_3 , AC_3 and TR_3 for the ROC surface are proposed and defined. And it is found that these five criteria for the ROC surface have relationship with the corresponding criteria for the ROC curves as well as the Kolmogorov-Smirnov statistics.

3. Three examples for the ROC surfaces

We consider three examples. The first and second examples were obtained from Scurfield (1996), and Nakas and Yiannoutsos (2004), respectively. And the final example was taken from other distributions. These examples are based on the normal distributions summarized in Table 3.1.

Table 3.1 Distribution functions of examples

Example	$F_1(\cdot)$	$F_2(\cdot)$	$F_3(\cdot)$
1.	$N(-1, 1)$	$N(0, 1.4^2)$	$N(1, 1.2^2)$
2.	$N(0, 1)$	$N(1, 1.1^2)$	$N(1.4, 1.3^2)$
3.	$N(0, 1)$	$N(1, 0.8^2)$	$N(2, 1.2^2)$

Table 3.2 Criteria for ROC curves

Example	Threshold	KS	J	SSS	MVD	AC	TR
1.	-0.2	0.3449	0.3449	1.3449	0.3449	0.6551	0.6725
	0.3	0.3050	0.3050	1.3050	0.3050	0.6816	0.6525
2.	0.6	0.3677	0.3677	1.3677	0.3677	0.6323	0.6838
	1.7	0.1465	0.1465	1.1465	0.1465	0.8535	0.5732
3.	0.4	0.4288	0.4288	1.4288	0.4288	0.5712	0.7144
	1.7	0.4079	0.4079	1.4079	0.4079	0.5921	0.7040

The results of Example 1 in Table 3.2 show that the values of KS , J , SSS , MVD , and TR for two distributions F_1 and F_2 were larger than those for F_2 and F_3 . On the other hand, AC for F_1 and F_2 had smaller value than for F_2 and F_3 , since AC attained the minimum value. Hence, the classification power for F_1 and F_2 was greater than for F_2 and F_3 . Examples 2 and 3 can be explained in a similar manner. Table 3.2 shows the optimal thresholds for F_1 and F_2 in Examples 1 to 3 were -0.2, 0.6 and 0.4. And the optimal thresholds for F_2 and F_3 in the three Examples were 0.3, 1.7 and 1.7. These thresholds were obtained based on the two dimensional ROC curves.

Table 3.3 Criteria for ROC surface

Example	Threshold	J_3	STR_3	MVD_3	AC_3	TR_3
1.	-0.2	0.1000	1.6499	0.6499	0.3501	0.5500
	0.3					
2.	0.6	0.0094	1.5142	0.5142	0.4858	0.5047
	1.7					
3.	0.4	0.2245	1.8367	0.8367	0.1633	0.6122
	1.7					

Table 3.3 shows three paired thresholds and values of the five optimal criteria proposed in Section 2 for the ROC surface. It is easy to find that the paired optimal thresholds in Table 3.3 are identical to those in Table 3.2. Also we found there are relationships between these values in Table 3.2 and 3.3 using the five corollaries in Section 2. Among five optimal criteria in Table 3.3, all values for Example 3 were larger than those for Examples 1 and 2. This means that the classification power for F_1 , F_2 and F_3 in Example 3 was greater than in Examples 1 and 2. This phenomenon could be derived from Figure 3.1.

Figure 3.1 demonstrates the three ROC surfaces corresponding to Example 1 to 3. From the ROC surfaces, we can understand the analytical interpretations discussed above. On each ROC surface, there is a mark which indicates the point corresponding to two optimal thresholds, $ROC_s(p_1, p_3) = \{F_2(F_3^{-1}(1-p_3)) - F_2(F_1^{-1}(p_1))\}$. The optimal threshold's mark of Example 3 locates higher and more closer to (1, 1, 1) point than other marks of Example 1 and 2. Therefore based on values of five optimal criteria and the corresponding marks on the ROC surface, the classification power of Example 3 is greater than other examples.

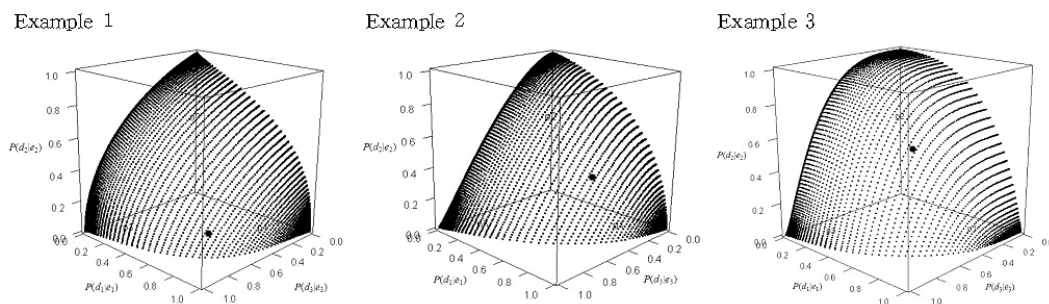


Figure 3.1 ROC surface

Table 3.4 contains three event–decision probability matrices of Example 1 to 3. For Example 1, the first and third true classification rates are high, and only the first true classification rate is high for Example 2. However, all three kinds of the true classification rates have higher values more than 0.5 in Example 3. Hence, we might conclude that Example 3 have the most classification power than Example 1 and 2 based on the results of Table 3.2, 3.3, 3.4 and Figure 3.1.

Table 3.4 Event-decision probability matrix

Decision	Ex 1 : Event			Ex 2 : Event			Ex 3 : Event		
	e_1	e_2	e_3	e_1	e_2	e_3	e_1	e_2	e_3
d_1	0.7881	0.4432	0.1587	0.7257	0.3581	0.2692	0.6554	0.2266	0.0912
d_2	0.1150	0.1416	0.1212	0.2297	0.3797	0.3221	0.3000	0.5826	0.3101
d_3	0.0968	0.4152	0.7202	0.0446	0.2623	0.4087	0.0446	0.1908	0.5987

4. Conclusion

The five classification criteria J , SSS , MVD , AC and TR for the two-dimensional ROC curve were extended to the three-class ROC surface. Since these criteria have a relationship to the Kolmogorov-Smirnov statistic, it was found that the criteria for the ROC surface can be represented by the paired Kolmogorov–Smirnov statistics to test both F_1 , F_2 and F_3 .

By using the five suggested classification criteria for the ROC surface, the paired optimal thresholds could be obtained. It was concluded that the paired optimal thresholds found from the ROC surface were identical to the two optimal thresholds selected from the two ROC curves.

In this paper, we assume only three events and decisions for a discrimination task. For more than three classes, the ROC analysis cannot be explained further. However for K (≥ 4) events and decisions, some optimal classification criteria might be extended and generalized. In particular, criteria such as KS , J , SSS , MVD , AC and TR could be developed for K -class classification problems. With these classification criteria, $K - 1$ optimal thresholds are obtained with ease, since the optimal thresholds from the ROC surface are equivalent to the optimal thresholds obtained from the ROC curves. Therefore, this work provides a framework for further research.

References

- Connell, F. A. and Koepsell, T. D. (1985). Measures of gain in certainty from a diagnostic test. *American Journal of Epidemiology*, **121**, 744-753.
- Dreiseitl, S., Ohno-Machado, L. and Binder, M. (2000). Comparing three-class diagnostic tests by three-way ROC analysis. *Medical Decision Making*, **20**, 323-331.
- Fawcett, T. (2003). *ROC graphs: notes and practical considerations for data mining researchers*, Technical report HPL-2003-4, HP Laboratories, 1-28.
- He, X. and Frey, E. C. (2006). Three-class ROC analysis - The equal error utility assumption and the optimality of three-class ROC surface using the ideal observer. *IEEE Transactions on Medical Imaging*, **25**, 979-986.
- Heckerling, P. S. (2001). Parametric three-way receiver operating characteristic surface analysis using mathematics. *Medical Decision Making*, **21**, 409-417.

- Hong, C. S. (2009). Optimal threshold from ROC and CAP curves. *Communications in Statistics-Simulation and Computation*, **38**, 2060-2072.
- Hong, C. S. and Joo, J. S. (2010). Optimal thresholds from non-normal mixture. *The Korean Journal of Applied Statistics*, **23**, 943-953.
- Hong, C. S., Joo, J. S. and Choi, J. S. (2010). Optimal thresholds from mixture distributions. *Journal of Korean Statistical Society*, **23**, 13-28.
- Krzanowski, W. J. and Hand, D. J. (2009). *ROC curves for continuous data*, Chapman and Hall, London.
- Mossman, D. (1999). Three-way ROCs. *Medical Decision Making*, **19**, 78-89.
- Nakas, C. T. and Yiannoutsos, C. T. (2004). Ordered multiple-class ROC analysis with continuous measurements. *Statistics in Medicine*, **23**, 3437-3449.
- Nakas, C. T., Alonzo, T. A. and Yiannoutsos, C. T. (2010). Accuracy and cut-off point selection in three-class classification problems using a generalization of the Youden index. *Statistics in Medicine*, **29**, 2946-2955.
- Nakas, C. T., Dalrymple-Alford, J. C., Anderson, T. J. and Alonzo, T. A. (2013). Generalization of Youden index for multiple-class classification problems applied to the assessment of externally validated cognition in Parkinson disease screening. *Statistics in Medicine*, **32**, 995-1003.
- Patel, A. C. and Markey, M. K. (2005). Comparison of three-class classification performance metrics: A case study in breast cancer CAD. *International Society for Optical Engineering*, **5749**, 581-589.
- Perkins, N. J. and Schisterman, E. F. (2006). The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American Journal of Epidemiology*, **163**, 670-675.
- Scurfield, B. K. (1996). Multiple-event forced-choice tasks in the theory of signal detectability. *Journal of Mathematical Psychology*, **40**, 253-269.
- Velez, D. R., White, B. C., Motsinger, A. A., Bush, W. S., Ritchie, M. D., Williams, S. M. and Moore, J. H. (2007). A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. *Genetic Epidemiology*, **31**, 306-315.
- Wandishin, M. S. and Mullen, S. J. (2009). Multiclass ROC analysis. *Weather and Forecasting*, **24**, 530-547.
- Yoo, H. S. and Hong, C. S. (2011). Optimal criterion of classification accuracy measures for normal mixture. *Communications for Statistical Applications and Methods*, **18**, 343-355.
- Youden, W. J. (1950). Index for rating diagnostic test. *Cancer*, **3**, 32-35.