Optimal thresholds criteria for ROC surfaces

C. S. Hong1 · E. S. Jung2

12Department 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}, \ldots, 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}, \ldots, X_{2n_2}\} and \{X_{31}, X_{32}, \ldots, 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:

\[
\begin{align*}
    \text{If } X &\leq c_1 \text{ then decision is class } 1 (d_1). \\
    \text{Else if } c_1 &\leq X < c_2 \text{ then decision is class } 2 (d_2). \\
    \text{Else decision is class } 3 (d_3).
\end{align*}
\]

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

\[
\begin{array}{cccc}
\text{Decision} & e_1 & e_2 & e_3 \\
\hline
d_1 & P(d_1|e_1) & P(d_1|e_2) & P(d_1|e_3) \\
d_2 & P(d_2|e_1) & P(d_2|e_2) & P(d_2|e_3) \\
d_3 & P(d_3|e_1) & P(d_3|e_2) & P(d_3|e_3) \\
\end{array}
\]

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

\[
\text{ROC}_s(p_1,p_3) = F_2(\frac{-1}{3}(1-p_3)) - F_2(\frac{-1}{1}(p_1)), \quad 0 \leq p_1, \ p_3 \leq 1,
\]

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): $TR = \max[F_1(x) + (1 - F_2(x))]$ (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

$$J_3 = \max_{x,y,x_y} \left[ \left\{ (F_1(x) - F_1(y) - F_1(x)) - (1 - F_1(y)) \right\} + \left\{ (F_2(y) - F_2(x)) - (F_2(x)) - (1 - F_2(y)) \right\} \right] + \left\{ (1 - F_3(y)) - F_3(x) - (F_3(y) - F_3(x)) \right\} / 3$$

$$= \max_{x,y,x_y} \left[ 2 \times F_1(x) + 2 \times (F_2(y) - F_2(x)) + 2 \times (1 - F_3(y)) \right] / 3.$$

**Corollary 2.1** Relation of $J$ and $KS$

$$J_3 = 2/3 \times (\max[J_{12}] + \max[J_{23}]) - 1/3$$

$$= 2/3 \times (KS_{12} + KS_{23}) - 1/3,$$

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 $KS$

$$STR_3 = SSS_{12} + SSS_{23} - 1$$

$$= KS_{23} + KS_{12} + 1,$$

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

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

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

**Corollary 2.3** Relation of $MVD$ and $KS$

$$MVD_3 = MVD_{12} + MVD_{23}$$

$$= KS_{23} + KS_{12},$$

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 $r_1$ and $r_2$ are obtained as

$$r_1 = \sqrt{\frac{(1 - F_1(x))^2 + (1 - (F_2(y) - F_2(x))^2 + (F_3(y))^2)}{1 - (F_1(x) - F_2(x) + F_2(y) - F_3(y))}},$$

$$r_2 = \sqrt{\frac{(1 - F_1(x))^2 + (1 - (F_2(y) - F_2(x))^2 + (F_3(y))^2)}{1 - (F_1(x) - F_2(x) + F_2(y) - F_3(y))}}.$$

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

$$AC_3 = \min[\sqrt{r_2^2/r_1^2}]$$

$$= \min[1 - (F_1(x) - F_2(x) + F_2(y) - F_3(y))].$$
Corollary 2.4 Relation of $AC$ and $KS$

\[ AC_3 = AC_{12} + AC_{23} - 1 \]
\[ = 1 - (KS_{12} + KS_{23}), \]

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

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

Corollary 2.5 Relation of $TR$ and $KS$

\[ TR_3 = \frac{2}{3}(TR_{23} + TR_{12} + 1/2) \]
\[ = \frac{1}{3}(KS_{12} + KS_{23} + 1), \]

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 Yiamoustos (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>
<thead>
<tr>
<th>Example</th>
<th>$F_1(\cdot)$</th>
<th>$F_2(\cdot)$</th>
<th>$F_3(\cdot)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>$N(-1,1)$</td>
<td>$N(0,1.4^2)$</td>
<td>$N(1,1.2^2)$</td>
</tr>
<tr>
<td>2.</td>
<td>$N(0,1)$</td>
<td>$N(1,1.1^2)$</td>
<td>$N(1,1.3^2)$</td>
</tr>
<tr>
<td>3.</td>
<td>$N(0,1)$</td>
<td>$N(1,0.8^2)$</td>
<td>$N(2,1.2^2)$</td>
</tr>
</tbody>
</table>

Table 3.2 Criteria for ROC curves

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

<table>
<thead>
<tr>
<th>Example</th>
<th>Threshold</th>
<th>$J_3$</th>
<th>$STR_3$</th>
<th>$MVD_3$</th>
<th>$AC_3$</th>
<th>$TR_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>-0.2</td>
<td>0.1000</td>
<td>1.6499</td>
<td>0.6499</td>
<td>0.3501</td>
<td>0.5500</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>0.6</td>
<td>0.0094</td>
<td>1.5142</td>
<td>0.5142</td>
<td>0.4858</td>
<td>0.5047</td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>0.4</td>
<td>0.2245</td>
<td>1.8367</td>
<td>0.8367</td>
<td>0.1633</td>
<td>0.6122</td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Table 3.4 Event-decision probability matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>d₁</td>
</tr>
<tr>
<td>d₂</td>
</tr>
<tr>
<td>d₃</td>
</tr>
</tbody>
</table>

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 $F_2$, $F_3$.

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 $K$ ($\geq 4$) 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


