2nd 2013 IEEE Conference on Control, Systems and Industrial Informatics (ICCSII)
Bandung, Indonesia, June 23-26, 2013

# Predicting Receiver Operating Characteristic Curve, Area Under Curve, and Arithmetic Means of Accuracies based on the Distribution of Data Samples

Ivanna Kristianti Timotius
Department of Electronic Engineering
Satya Wacana Christian University
Salatiga, Indonesia
ivanna.timotius@ieee.org

***Abstract*—Measuring the performance of a classifier is an essential step in building a classification method for a two class classification problem. The Receiver Operating Characteristic (ROC) Curve, Area Under ROC Curve (AUC), and Arithmetic Means of Accuracies (Ameans) are several classifier performance measurements that are typically calculated by conducting an experiment. This paper presents predicting methods of these classifier performance measurements based on the data sample distributions. The experiment shows that the predicting methods results are similar with the empirical results using the testing data set. Therefore the methods are applicable in predicting the classifier performance without conducting an experiment. The predicted performance measurements might be useful in evaluating the discriminability of a feature sample.**

***Keywords*—classifier performance measurement; receiver operating characteristic curve; area under curve; arithmetic means of accuracies; data sample distribution***

## I. INTRODUCTION

The performances of some classifiers depend on the arbitrary selection of decision thresholds. In such classifiers, the Receiver Operating Characteristic (ROC) curve is employed to give an empirical description of a decision threshold effect [1]. In the case of two class classification based on a decision threshold, if the data distributions of the positive and negative data samples are depicted in Fig. 1, the performance of the classifier is greatly depend on the chosen decision threshold. Moving the decision threshold to the negative class will decrease the number of correctly classified data on the negative class, but will increase the number of correctly classified data on the positive class. Changing the decision threshold to any value will give different numbers of correctly classified data on the negative and positive classes. An ROC curve is used to depict the effect of changing this decision threshold.

The ROC curve is typically shaped empirically by using a testing data set. However, this paper proposes a method to predict the ROC curve based on the distribution of the data samples. Knowing the predicted ROC curve, the predicted Area Under ROC Curve (AUC) can be calculated as well. In addition, this paper also proposes an algorithm to predict the value of decision threshold having the highest value of arithmetic means of accuracies (Ameans) [2] based on the distributions of the data samples.

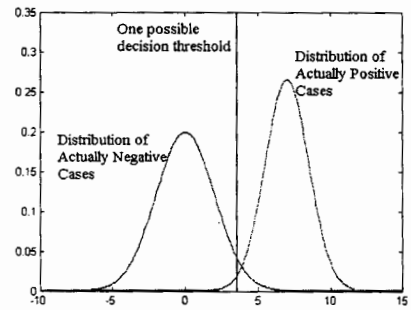


Figure 1. An example of the distribution of data samples

This predicted performance measurements might give information about the discriminability of a feature fed to a classifier. Commonly, a discriminative feature is a feature having high mean difference between classes and low variation within classes. If a non-discriminative feature is fed to a classifier, it might decrease the classification accuracies and/or increase the classification time. The authors in [3] define the discriminability (or relevance) of a feature based on the mean difference between the classes. This paper proposes a performance measurement prediction method that might be used to evaluate the discriminability of a feature based on the average and the standard deviation of the feature samples (assumed that the feature sample distributions are Gaussian) or based on the distribution of the feature sample.

The rest of this paper is organized as follows. Section 2 gives a brief explanation of ROC curve and AUC. Section 3 reviews the concept of Ameans. Section 4 presents the ROC, AUC, and Ameans predicting methods. Section 5 presents the experimental design and results of the methods. Finally, Section 6 presents the conclusions of the work.

## II. RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE AND AREA UNDER ROC CURVE (AUC)

An ROC curve or an ROC graph [4] is a depiction of classifier performance in an ROC space. An ROC space is a two-dimensional space in which the true positive rate (TP rate)

978-1-4673-5817-0/13/$31.00 ©2013 IEEE

is plotted on the Y axis and the false positive rate (FP rate) is plotted on the X axis [1][5]. The TP rate also known as true positive fraction (TPF) or sensitivity is defined as:

$$\text{TP rate} = \frac{\text{number of positives correctly classified}}{\text{number of total positives}} \quad (1)$$

Whereas, the FP rate also known as the false positive fraction (FPF) is defined as:

$$\text{FP rate} = \frac{\text{number of negatives incorrectly classified}}{\text{number of total negatives}} \quad (2)$$

In a two-class classifier using a decision threshold, each value of decision threshold will give a pair of TP rate and FP rate, then consequently will give a point in the ROC space. By conducting experiments using several decision thresholds, the classifier will produce several points in the ROC space. These points can be connected to produce a curve which is called an ROC curve.

The point (0, 0) in ROC space represents the never issuing a positive classification strategy. The point (1, 1) in ROC space represents the strategy of always issuing positive classification. The point (0, 1) represents a perfect classification. Basically, it is desirable to have a classifier with a high TP rate and low FP rate.

The Area Under ROC Curve (AUC) is one possible method to compare classifiers based on their ROC curves [4][6]. Each ROC curve produces one value of AUC. The range of AUC is between 0 and 1 since AUC is a portion of ROC space having an area of a unit square.

## III. Arithmetic Means of Accuracies

The arithmetic means of accuracies (Ameans) is one possible method in comparing classifiers between several points in the ROC space [2]. A non-weighted Ameans for a two-class classification is defined as:

$$\text{Ameans} = \frac{1}{2}(\text{TP rate} + \text{TN rate}) \quad (3)$$

where

$$\text{TN rate} = 1 - \text{FP rate} = \frac{\text{number of negatives correctly classified}}{\text{number of total negatives}} \quad (4)$$

The value of Ameans is ranging between 0 and 1. Ameans can be used to measure the performance of classifiers that can only create one point in the ROC space. A better classifier is a classifier producing a higher Ameans.

## IV. Predicting the ROC Curve, Area Under ROC Curve, and the Arithmetic Means of Accuracies

The ROC curve predicting method proposed in this paper is based on the probability density functions of the negative data sample $f_N(x)$ and the probability density function of positive data sample $f_P(x)$. The cumulative distribution functions for the negative and positive data samples are respectively denoted by $F_N(x)$ and $F_P(x)$. Note that if a classifier classifies the data samples based on a feature of the data, $f_N(x)$ and $f_P(x)$ are the probability density function of the feature samples. Also if a classifier is based on a parameter calculated from the feature samples, $f_N(x)$ and $f_P(x)$ are the probability density function of the parameter. Given a decision threshold, $t_h$, the predicted TN rate, FP rate, TP rate, and FN rate are calculated by the following equations.

$$\text{Predicted TN rate} = F_N(t_h) \quad (5)$$

$$\text{Predicted FP rate} = 1 - \text{Predicted TN rate} \quad (6)$$

$$\text{Predicted FN rate} = F_P(t_h) \quad (7)$$

$$\text{Predicted TP rate} = 1 - \text{Predicted FN rate} \quad (8)$$

By changing the decision threshold, $t_h$, several points in the ROC space can be generated. Thus, an ROC curve can be predicted. By knowing the ROC curve, the predicted AUC of the ROC curve can be calculated. The Ameans of each point in the ROC curve and the maximum value of the Ameans can be calculated as well.

One possible method to choose the best decision threshold in a classifier is by observing the point in ROC curve which has the maximum value of Ameans. This paper proposes a method of choosing this decision threshold, $t_{best}$, so that the maximum value of Ameans can be calculated by the following equation.

$$\text{Predicted Max Ameans} = \frac{1}{2}\left(F_N(t_{best}) + (1 - F_P(t_{best}))\right) \quad (9)$$

The method of choosing the decision threshold, $t_{best}$, starts with predicting that the highest Ameans is obtained if the $t_{best}$ is set at the intersection of the negative and positive probability density functions as shown in the following equation.

$$f_N(t_{best}) = f_P(t_{best}) \quad (10)$$

In the case that $f_N(x)$ is a Gaussian probability density function [7] having mean $m_N$ and standard deviation $\sigma_N$, and $f_P(x)$ is a Gaussian probability density function having mean $m_P$ and standard deviation $\sigma_P$, the best decision threshold, $t_{best}$, can be calculated using the following equation.

$$t_{best} = \begin{cases} \dfrac{m_P^2 - m_N^2}{2(m_P - m_N)}, & \text{if } \sigma_N = \sigma_P \\ \dfrac{-b_1 + \sqrt{b_1^2 - 4a_1c_1}}{2a_1}, & \text{if } \sigma_N < \sigma_P \\ \dfrac{-b_2 + \sqrt{b_2^2 - 4a_2c_2}}{2a_2}, & \text{if } \sigma_N > \sigma_P \end{cases} \quad (11)$$

where

$$a_1 = \sigma_P^2 - \sigma_N^2 \quad (12)$$

$$b_1 = 2(m_P\sigma_N^2 - m_N\sigma_P^2) \quad (13)$$

$$c_1 = m_N^2\sigma_P^2 - m_P^2\sigma_N^2 - 2\sigma_N^2\sigma_P^2 \ln\left(\frac{\sigma_P}{\sigma_N}\right) \quad (14)$$

$$a_2 = \sigma_N^2 - \sigma_P^2 \quad (15)$$

$$b_2 = 2(m_N\sigma_P^2 - m_P\sigma_N^2) \quad (16)$$

$$c_2 = m_P^2\sigma_N^2 - m_N^2\sigma_P^2 - 2\sigma_N^2\sigma_P^2 \ln\left(\frac{\sigma_N}{\sigma_P}\right) \quad (17)$$

## V. EXPERIMENTS AND RESULTS

The experiments are conducted by generating Gaussian distributed random data. The class distributions have $m_N = 4$, $n_P = 8$, $\sigma_N = 3$, and $\sigma_P = 2$. The first experiment generates 1000 data for each class. The second experiment used an imbalanced data set (1000 data for the positive class and 500 data for the negative class). By changing the decision threshold, the empirical ROC curves of this classifier are shown in Fig. 2. The AUCs calculated from the empirical ROC curves are shown in Table I. In the experiment by using these random data, the empirical Ameans obtained by the decision thresholds are calculated. Subsequently, the point having the maximum values of Ameans are shown in Fig. 2.

The predicted ROC curve calculated based on the mean and the standard deviation of the classes is shown in Fig. 2. The predicted point with the maximum value of Ameans which is obtained from the predicting method is also shown in Fig. 2. The value of these AUC and Ameans of Fig. 2 are shown in Table I.

The predicted ROC curve shown in Fig. 2 is similar with the ROC curves obtained empirically from the balanced data set and the imbalanced data set. The predicted AUC and predicted maximum Ameans is similar with the experiment results as well. Therefore, these prediction methods are applicable in predicting the classifier performances.

TABLE I. THE AUC AND AMEANS OF FIG. 2

| Classifier Performance Measurement | Calculation Methods | | |
|---|---|---|---|
| | *Experiments* | | *Prediction* |
| | *Balanced* | *Imbalanced* | |
| AUC | 87.35% | 87.12% | 86.63% |
| Maximum value of Ameans | 80.00% | 79.90% | 79.50% |

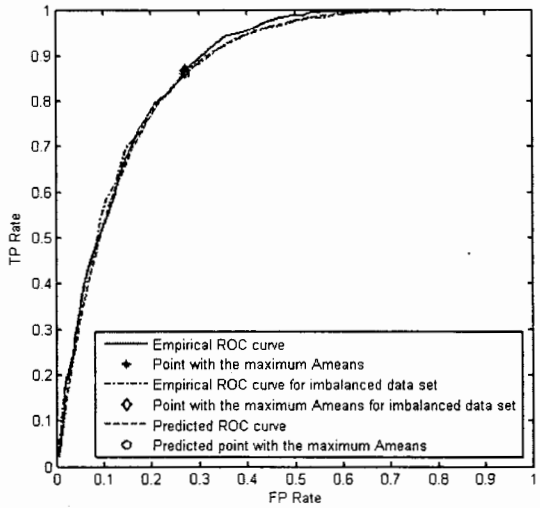


Figure 2. The predicted ROC Curve

The effects of the data sample mean and standard deviation to the predicted ROC curve, AUC, and maximum Ameans are shown in Fig. 3 and Fig. 4. The circles shown in Fig. 3 and Fig. 4 depict the points in ROC space having the maximum Ameans at the associated ROC curves. The intersection appeared in Fig. 4 indicates that there is a point with equal performace resulted from different values of standard deviation and different values of decision threshold, $t_h$.

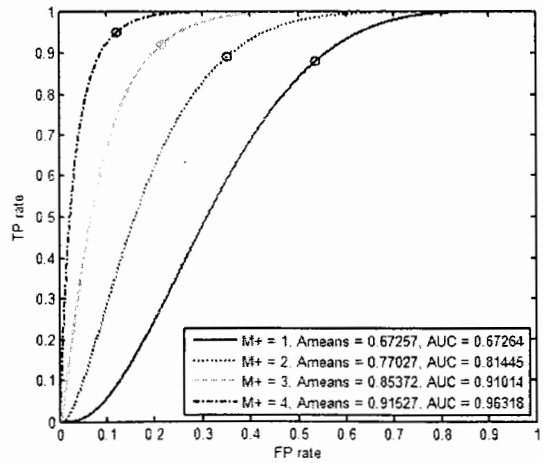


Figure 3. Predicted ROC curves with $m_N = 0$, $\sigma_N = 2$, and $\sigma_P = 1$.

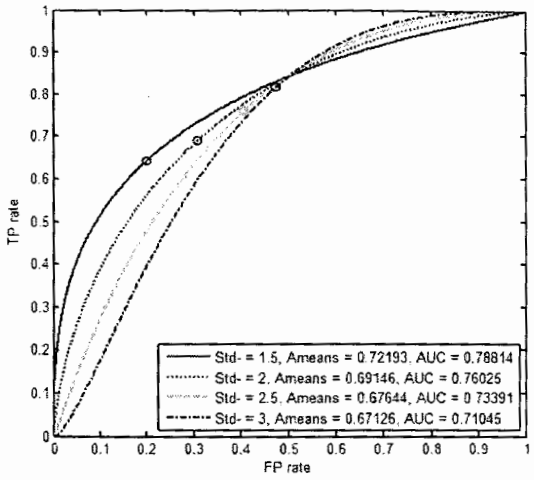


Figure 4. Predicted ROC curves with $m_N = 0$, $m_P = 2$, and $\sigma_P = 2$.

The ROC curve, AUC, and Ameans significantly depend on the mean and standard deviation of the data samples. Commonly, the data samples with higher mean difference and lower standard deviations are easier to be classified.

The next experiments are done using the second and third class of Iris flower data set [8]. Each class of the Iris data set contains 50 data. The data set consists of information about the sepal length as the first feature, sepal width as the second feature, petal length as the third feature, and petal width as the fourth feature of the flowers. The average and the standard deviation of the second and third class of this Iris data set are given in Tabel II. Using these statistical values and an assumsion that the data distributions are Gaussian, the ROC curves, AUC, and maximum value of Ameans from classifiers based on decision thresholds are predicted and shown in Table IV and Fig. 5. In Fig. 5 the predicted ROC curves from the first feature until the fourth feature are respectively denoted by ROC1, ROC2, ROC3, and ROC4. The predicted points with the maximum value of Ameans assosiated with the ROC curves are respectively denoted by Ameans1, Ameans2, Ameans3, and Ameans4 in Fig. 5.

Further experiments are done using the Euclidean distance between the data and the average of the second and third class. The average and standard deviation of the Euclidean distances are shown in Table III. The ROC curves, AUC, and maximum value of Ameans based on these data are predicted and shown in Table IV and Fig. 5. In Fig. 5 the predicted ROC curves obtained from the experiment based on the data sample distance to the average of the second and third class are respectively denoted by ROCc2 and ROCc3. The predicted points having the maximum value of Ameans assosiated with the ROC curves are denoted respectively by Ameansc2 and Ameansc3 in Fig. 5.

The results in Table IV and Fig. 5 give us information about the discriminability of the features. The highest discriminability is obtained by using the fourth feature of the Iris data set which has the lowest standard deviation within the classes. The second best discriminability is obtained using the third feature which has the highest average difference between the classes.

TABLE II. THE STATISTICAL VALUE OF THE SECOND AND THIRD CLASS IN IRIS DATA SET

| Statistical Value | First Feature | Second Feature | Third Feature | Fourth Feature |
|---|---|---|---|---|
| Average of the 2nd Class | 5.94 | 2.77 | 4.26 | 1.33 |
| Average of the 3rd Class | 6.59 | 2.97 | 5.55 | 2.03 |
| Average difference of the 2nd and 3rd class | 0.65 | 0.20 | 1.29 | 0.70 |
| Standard deviation of the 2nd class | 0.52 | 0.31 | 0.47 | 0.20 |
| Standard deviation of the 3rd class | 0.64 | 0.32 | 0.55 | 0.27 |

TABLE III. THE STATISTICAL VALUE OF THE DISTANCE BETWEEN THE CLASSES AND THE AVERAGE OF THE CLASSES

| Distance | Average | Standard Deviation |
|---|---|---|
| 2nd class to the average of 2nd class | 0.71 | 0.34 |
| 3rd class to the average of 2nd class | 1.74 | 0.70 |
| 2nd class to the average of 3rd class | 1.70 | 0.60 |
| 3rd class to the average of 3rd class | 0.82 | 0.45 |

TABLE IV. THE PREDICTED AUC AND AMEANS OF FIG. 5

| Feature of Iris data set | Performance Measurement | |
|---|---|---|
| | *AUC* | *Maximum value of Ameans* |
| First feature (sepal length) | 78.47% | 71.75% |
| Second feature (sepal width) | 67.35% | 62.59% |
| Third feature (petal length) | 96.22% | 89.74% |
| Fourth feature (petal width) | 97.93% | 93.20% |
| Distance to the mean of the 2nd class | 90.86% | 85.38% |
| Distance to the mean of the 3rd class | 87.83% | 80.13% |

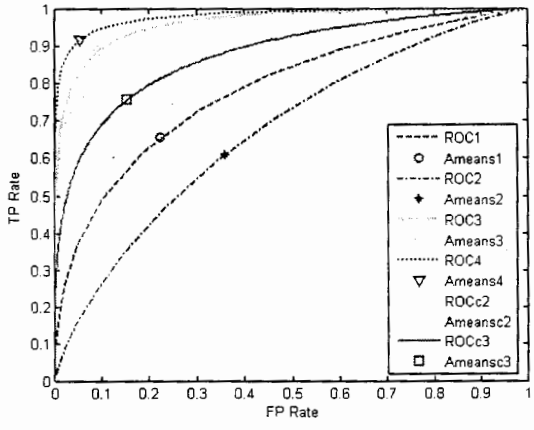


Figure 5. Experimental results using the Iris data set.

## VI. CONCLUSIONS

Methods in predicting the ROC curve, AUC, and the maximum Ameans based on the sample distributions have been presented. These methods are shown having the similar results with the experimental results.

The predicted performace measurements might be useful as tools for feature sample discriminability examination. This is for the reason that the measurements depend on not simply by the mean difference between classes, but also the standard deviation within classes (if the feature samples are assumed to have Gaussian distributions). In the future, we will conduct further research on the feature discriminability examination.

## REFERENCES


[1] C. E. Metz, "Basic principle of ROC analysis," Seminars in Nuclear Medicine, vol.VIII, no. 4, 1978.

[2] I. K. Timotius and S. G. Miaou, "Arithmetic means of accuracies: a classifier performance measurement for imbalanced data set," Proc. of Int. Conf. on Audio, Language, and Image Processing, 2010.

[3] M. T. Coimbra and J. P. S. Cunha, "MPEG-7 visual descriptor – contribution for automated feature extractor in capsule endoscopy," IEEE Trans. on Circuit and Systems for Video Technology, vol. 16, no. 5, pp. 628 – 637, 2006.

[4] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, issue 8, pp. 861 – 874, 2006.

[5] F. Provost and T. Fawcett, "Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions," Proc. of the Third Int. Conf. in Knowledge Discovery and Data Mining, pp. 43 – 48, 1997.

[6] C. X. Ling, J. Huang, and H. Zhang, "AUC: a statistically consistent and more discriminating measure than accuracy," Proc. of Int Joint Conf. on Artificial Intelligence, 2003.

[7] P. Z. Peebles, Probability, Random Variables, and Random Signal Principles, 3rd ed., McGraw-Hill: Singapore, 1993.

[8] R. A. Fisher, "The use of multiple measurements in taxonomic problems," Annals of Eugenics, vol. 7, no 2, pp. 179–188, 1936.