On ROC Curve Analysis of Artificial Neural Network Classifiers

Chulwoo Kim, Sung-Hyuk Cha
Computer Science Department, Pace University
1 Pace Plaza
New York, NY 10038

Yoo Jung An, Ned Wilson
Essex County College
303 University Ave
Newark, NJ 07102

Abstract

Receiver operating characteristic or ROC curves are of great interest in evaluating many security systems such as biometric authentication. They visualize the trade-off between the number of security breaches and the level of convenience. In the earlier work, ROC curves and their decision boundaries were studied for various classifiers. Here, further studies are conducted to identify problems of ROC curve analysis when artificial neural network (ANN) classifiers’ net values are used. Graphical decision boundaries and experimental results on the IRIS biometric authentication system reveal the over-fitting in the ROC curve analysis. This graphical decision boundaries suggest that ANN classifiers with two output units are more desirable than those with a single output unit for two class classification problems. Some problems are identified in some conventional ROC curve analysis tools in ANN with multiple output units and suggest a suitable model.

Introduction

Receiver operating characteristic or simply ROC curve was first used to analyze radar signals (Green and Swets 1966) and has been employed in machine learning and pattern recognition areas to evaluate classification algorithms (Fawcett 2006; Bradley 1997; Duda, Hart, and Stork 2012). It is of great importance not only in cyber security (Adams and Heard 2014), but also in many other fields such as computer vision (Jones and Rehg 2002) and medical science (Cook 2008).

In a biometric authentication system, two types of errors are introduced: false negative rate (FNR) and false positive rate (FPR). The ROC curve shows the relationship between these errors. The ROC curve can be obtained trivially by altering the scalar threshold value in a simple match model for a biometric authentication system. While most parametric and non-parametric such as support vector machine result smooth and reasonable decision boundaries as shown in Figure 1 (a) and (b), respectively, artificial neural networks showed astoundingly strange decision boundaries as shown in Figure 1 (c) ~ (d) (see (Cha, An, and Tappert 2010) for various parametric and non-parametric classifiers’ decision boundary figures).

Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
A simple distance based biometric match model utilizes a certain distance measure between two biometric data and the scalar distance value is classified based on the threshold value \( t \) as defined in (3) on the belief that the within-class distance tends to be smaller than the between-class distance.

\[
c(x, y) = \begin{cases} 
  w & \text{if } d(x, y) \leq t \\
  b & \text{otherwise}
\end{cases}
\]  

For a fixed value \( t \), two types of error probabilities can be determined. First, False Negative Rate (FNR) is the probability of within-class data classified as between-class as given in (4). Next, False Positive Rate (FPR) is that of between-class data classified as within-class as in (5). True Positive Rate (TPR) and True Negative Rate (TNR) which are correct cases are defined in (6) and (7), respectively.

\[
FNR = Pr(c(x, y) = b/s(x) = s(y)) \quad \text{(4)}
\]
\[
FPR = Pr(c(x, y) = w/s(x) \neq s(y)) \quad \text{(5)}
\]
\[
TPR = Pr(c(x, y) = w/s(x) = s(y)) \quad \text{(6)}
\]
\[
TNR = Pr(c(x, y) = b/s(x) \neq s(y)) \quad \text{(7)}
\]

By changing the threshold value \( t \) in Figure 2 (a), the typical ROC curve can be obtained as shown in Figure 2 (b). FPR is often referred to as False Accept Rate (FAR), False Match Rate (FMR), Type I error, or simply a false alarm. FNR is often called the False Reject Rate (FRR), False Non-Match Rate (FNMR), Type II error, or simply a miss.

In order to visualize the decision boundary, we consider hypothetical two dimensional data samples. If two samples come from a same person and measure each feature distance, it becomes a point in feature distance space and it belongs to within class \( w \). If two samples are from two different people, it belongs to between class \( b \). This transformation from feature space to feature distance space was called dichotomy transformation (Cha and Srihari 2000). This Dichotomy model was first introduced in (Cha and Srihari 2000) to assess the power of individuality of handwriting where various pattern classification algorithms can be applied.

ROC curve analysis is one of the popular methods to evaluate various classifiers’ performance. Judging from ROC curves in Figure 3 alone, the ANN classifier seems to be better than other classifiers. Yet, hypothetical decision boundary analysis suggests that ANN classifiers have very complex boundaries over the threshold \( t \) whereas others have reasonable smooth boundaries.

**ROC curves for Artificial Neural Networks**

The Artificial Neural Network (ANN) has been widely utilized to solve classification problems (Duda, Hart, and Stork 2012; Mitchell 1997). A typical ANN classifier is consisted of input, hidden, and output layers of neurons. Feed forward artificial neural network classifiers can be learned using training sets. Samples in within class \( W \) are trained to be 1 and samples in between class \( B \) are trained to be 0 for an artificial neural network with a single output neuron. Given input values \((x, y)\), let \( net(x, y) \) be the net output value. Then the predicted decision can be made using (8).

\[
c(x, y) = \begin{cases} 
  w & \text{if } net(x, y) \geq t \\
  b & \text{otherwise}
\end{cases}
\]  

A typical threshold value \( t \) would be 0.5 but if \( t \) changes, an ROC curve can be generated. Decision boundaries when the threshold \( t \) changes reveal are very complex as illustrated in Figure 4. While the ROC curve demonstrates excellent results on the training set according to the ROC curve analysis, dramatic differences are observed in testing sets. Each time an ANN with same structure results in dramatically different decision boundaries. The net values may not be suitable for ROC curve analysis.

The ROC curve for multiple class classification problems had been studied such as in (Hand and Till 2001). Conventionally, there are \( c \) number of output units in the output
layer for the e number of class classification problem. Let net\(_x\)(X) denote the net value of output neuron of the class x where x ∈ C where C is the set of all classes. Prediction decision is made by the output neuron which fires the highest value as given in (9).

\[
c(X) = \arg\max_{x \in \text{outs}} \text{net}_x(X)
\]  

(9)

So as to utilize an ROC curve analysis, a different prediction decision rule with a threshold value \(t\) over their net values such as in (10) is necessary.

\[
c(X) = \begin{cases} 
\arg\max_{x \in \text{outs}} \text{net}_x(X) & \text{if } \max_{x \in \text{outs}} \text{net}_x(X) \geq t \\
\text{void} & \text{otherwise}
\end{cases}
\]

(10)

Here no decision is made if maximum net value is less than the threshold. A dichotomic ANN classifier can be designed with two output units instead of a single output unit. While a typical ROC curve is generated from 2 × 2 contingency table in an ANN with a single output unit, the 2 × 3 contingency table given in Table 1 is used with additional void positive rate (VPR) and void negative rate (VNR). Let \(t(X)\) and \(c(X)\) be the actual truth class and the predicted class by the classifier, respectively. There are six rates in 2 × 3 contingency table and defined in (11) ~ (16).

\[
\text{FPR} = \Pr((c(X) = b \land \text{net}_b \geq t)/(t(X) = w))
\]  

(11)

\[
\text{FNR} = \Pr((c(X) = b \land \text{net}_b \geq t)/(t(X) = w))
\]  

(12)

\[
\text{TPR} = \Pr((c(X) = w \land \text{net}_w \geq t)/(t(X) = b))
\]  

(13)

\[
\text{TNR} = \Pr((c(X) = b \land \text{net}_b \geq t)/(t(X) = w))
\]  

(14)

\[
\text{VPR} = \Pr((\text{net}_w < t \land \text{net}_b \geq t)/(t(X) = w))
\]  

(15)

\[
\text{VNR} = \Pr((\text{net}_w < t \land \text{net}_b < t)/(t(X) = b))
\]  

(16)

Artificial neural networks with two output units following the classification rules in (11) ~ (16) result smooth decision boundaries when \(t\) changes as shown in Figure 5. Although there are regions with no decision, no strange decision boundaries are observed. When designing an ANN for a two class classification problem, two output unit version seem to perform better than one output unit version in terms of ROC curve analysis.

Figure 6 (a) and (b) show the ROC curves for FPR vs. TPR and FNR vs. TNR for a 16-10-2 ANN IRIS biometric authentication classifier using a commercial neural network software (see MathWorks 2016). The blue sold and red dashed lines represent the curves for training and testing sets, respectively. There are some flaws in these ROC curve plots.

In order to understand the flaws, ten different scenarios for net outputs as depicted in Figure 7 must be considered where blue and red bars represent the within class net value and the between class net value, respectively. The first and second column cases in the table in Figure 7 correspond to the conventional 2 × 2 contingency table. The last column is void case where both net output values are below the threshold and thus no decision can be made. Third and fourth cases can be classified according the decision rules in (11) ~ (14). However, the conventional ROC curve plotting softwares such as (MathWorks 2016) are plotted by the following decision rules in (20) ~ (19).

\[
\text{TPR} = \Pr((\text{net}_w \geq t)/(t(X) = w))
\]  

(17)

\[
\text{FPR} = \Pr((\text{net}_b < t)/(t(X) = b))
\]  

(18)

\[
\text{TNR} = \Pr((\text{net}_b \geq t)/(t(X) = b))
\]  

(19)

\[
\text{FNR} = \Pr((\text{net}_w < t)/(t(X) = w))
\]  

(20)

Several anomalies are discovered.

**Anomaly 1. TN vs. FP anomaly**
If \(t(X) = b\) and \(\text{net}_w > \text{net}_b > t\), it falls into TN according the decision rule in (19) while it falls into FP according to the decision rule in (12).

**Anomaly 2. TP vs. FN anomaly**
If \(t(X) = w\) and \(\text{net}_b > \text{net}_w > t\), it falls into TP according to the decision rule in (17) while it falls into FN according to the decision rule in (11).

Intuitively, decisions rules in (11) ~ (16) make more intuitive senses than those in (17) ~ (20). Yet, while ROC curves based on decision rules in (17) ~ (20) provide smooth ones as given in Figure 6 (a) and (b), ROC curves based on decisions rules in (11) ~ (16) did not provide smooth curves as given in Figure 6 (c) and (d). This indicates that the anomaly cases are quite abundant in ROC curve analysis.
The iconic feed forward neural network for XOR problem (Werbos 1974) has a single output neuron. It is questionable whether one should use a single or two output units in a feed forward artificial neural network for two class classification problem. It was shown experimentally and with visual graphics that one with two output units is better than one with a single output unit in terms of ROC curve analysis. When the threshold value changes, some reasonable decision boundaries were derived in ANNs with two output neurons whereas strange decision boundaries were observed in those with a single output unit.

Since ROC curve analyses for many applications with artificial neural networks are used pervasively, studies to find meaning of ROC curves especially for ANNs were conducted in this article. When two or more output units are used in ANN, void cases happen and those are not used in ROC curve analysis and decisions cannot be made if all output unit net values are below the threshold value. This article reviewed some commercial ROC curve analysis products and identified some anomalies. Further studies are necessary to utilize the ROC curve analysis in artificial neural network classifiers.

**References**


