Receiver Operating Characteristic (ROC) Analysis

Nicolas Lachiche
University of Strasbourg, France

INTRODUCTION

Receiver Operating Characteristic (ROC curves) have been used for years in decision making from signals, such as radar or radiology. Basically they plot the hit rate versus the false alarm rate. They were introduced recently in data mining and machine learning to take into account different misclassification costs, or to deal with skewed class distributions. In particular they help to adapt the extracted model when the training set characteristics differ from the evaluation data. Overall they provide a convenient way to compare classifiers, but also an unexpected way to build better classifiers.

BACKGROUND

ROC analysis mainly deals with binary classifiers, models associating each entry to the positive or to the negative class. The performance of a given classifier is collected in a confusion matrix (also known as a contingency table) counting the number of training examples in each of the four cells depending on their actual classes and on their predicted classes (see Table 1).

The True Positive Rate (TPR) is the fraction of positive examples correctly classified, TP/P and the False Positive Rate (FPR) is the fraction of negative examples incorrectly classified, FP/N.

MAIN FOCUS

ROC Space

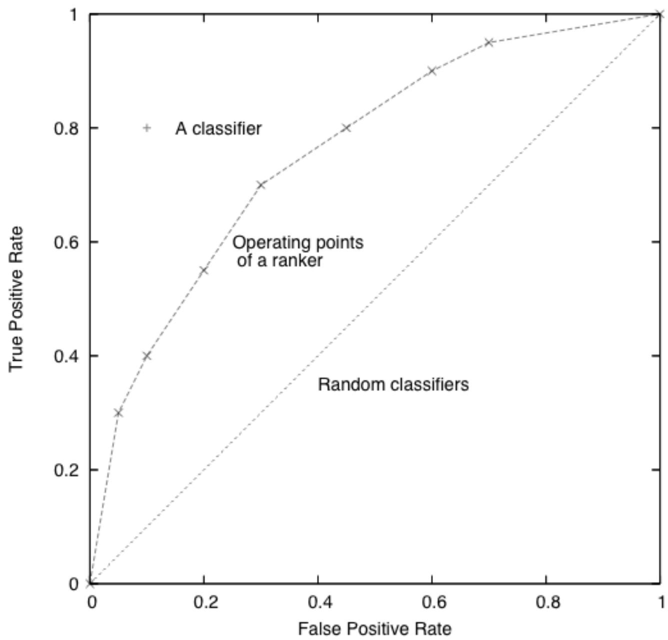
A ROC curve plots the True Positive Rate (TPR, aka recall or sensitivity) versus the False Positive Rate (FPR, equals to 1 – specificity). The performance of a binary classifier corresponds to a single point in the ROC space, cf. A classifier on Figure 1.

A ranker is a model associating a score to each entry, e.g. the naive bayesian classifier. Intuitively this score represents the "probability" that the entry is positive, even if it is not always a proper probability. Overall it provides a ranking of all examples. A threshold can be chosen and all examples whose score is above (respectively below) that threshold are predicted as positive (resp. negative). This is a called an Operating Point. It turns a ranker into a classifier. Different operating points lead to different classifiers. Therefore a ranker leads to a whole ROC curve, cf. Operating points of a ranker on Figure 1. Actually the curve is not continuous, only the operating points make sense. The number of operating points is the number of different ways of splitting the ranking of the examples. It is the number of examples if there are no tie, less otherwise. An efficient algorithm (Fawcett, 2004) to build the ROC curve is:

Table 1.

	Real Positive	Real Negative
Predicted Positive	True Positive (TP) aka Hit	False Positive (FP) aka False Alarm aka Type I Error
Predicted Negative	False Negative (FN) aka Miss aka Type II Error	True Negative (TN) aka Correct Rejection
	Total Number of Positive (P)	Total Number of Negative (N)

Figure 1. A classifier and a ranker in the ROC space



1. start in (0,0) with a score equals to the infinity
2. for each example ordered by decreasing scores
    1. if its score is different from the current score, plot a point in the current position and store its score as the current one
    2. if the example is positive, move up by 1/P, if it is negative move by 1/N

The ROC space contains several well-known landmarks:

- the top left corner corresponds to a perfect classifier, predicting correctly all positive and all negative examples,
- the bottom left corresponds to the constantly negative classifier, always predicting negative,
- the top right corner corresponds to the constantly positive classifier,
- the bottom-left to top-right diagonal denotes the performances of random classifiers, indeed any point can be obtained by a classifier randomly predicting positive a constant proportion of the entries, e.g. (0.3,0.3) corresponds to a classifier randomly predicting positive 30% of the entries.

Selection Of The Best Model According To The Operating Conditions Independently Of The Training Conditions

A ROC curve is first useful to select an operating point for a ranker or to select one among different classifiers. It should be noted that a ROC curve is insensitive to the class distribution because TPR and FPR do not depend on it. Whatever the class distribution and misclassification costs were on the training set, it is possible to select the best model according to their values on the new population to predict. Indeed once the expected number of positive P', the expected number of negative N', and the cost of misclassifying a positive (resp. negative) example C(-/+) (resp. C(+/-)) are known, it is easy to select the classifier:

- either maximizing the accuracy = P' * TPR + N' * (1 – FPR)
- or minimizing the cost = C(+/-) * FPR + C(-/+) * (1 - TPR).