Option #1: ROC vs. Lift Chart

Scott Miner

Colorado State University – Global Campus

**Option #1: ROC vs. Lift Chart**

The assignment asks the student to compare Receiver Operating Characteristic (ROC) curves and Lift Charts. ROC curves plot the true positive rate (TPR) of a classifier by its false positive rate (FPR) (Sinha & May, 2004).We express TPR as The equation for FPR is . *T*, *F*, *P*, and *N* stand for *True*, *False*, *Positive*, and *Negative*, respectively. We compute these measures from confusion matrices. Figure 1 displays a confusion matrix and the corresponding measures calculated from it. ROC curves graph the TPR of the classifier along the *y*-axis and the FPR along the *x*-axis. ROC curves depict tradeoffs between benefits and costs. Figure 2 shows a ROC graph with five discrete classifiers labeled A through E. Discrete classifiers output a single confusion matrix for a given model, the TPR and FPR of which we plot in the ROC space. In figure 2, the lower left point (0,0) represents the strategy of never making a positive classification. The point in the upper-righthand corner (1,1) represents the strategy of making all positive classifications. The diagonal line through the middle represents the strategy of randomly deducing a class. For example, a classifier that unsystematically guesses the positive group half the time is expected to get half the positive and half the negative forecasts correct, yielding a point (0.5,0.5) along the diagonal line in the ROC space. A classifier that unsystematically guesses the positive class 90% of the time, on the other hand, produces a point (0.9,0.9) in the ROC space and increases the FPR to 90%. In Figure 2, the point (0,1), D, represents a model with perfect classification. Model E performs worse than random assignment and is the negation of B. C’s performance is near-random, correctly guessing the positive class 70% of the time. Model B is more liberal than A, meaning it classifies more positive records correctly at the cost of classifying more negative instances incorrectly (Fawcett, 2006).

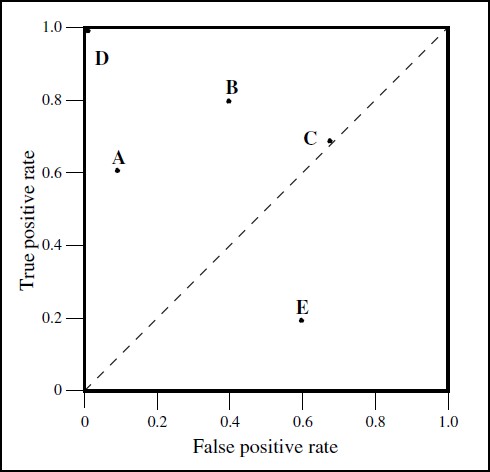


Figure 2. Basic ROC curve with five discrete classifiers. Reprinted from “An introduction to ROC analysis," by T. Fawcett, 2006, Pattern Recognition Letters, 27(8), 862.

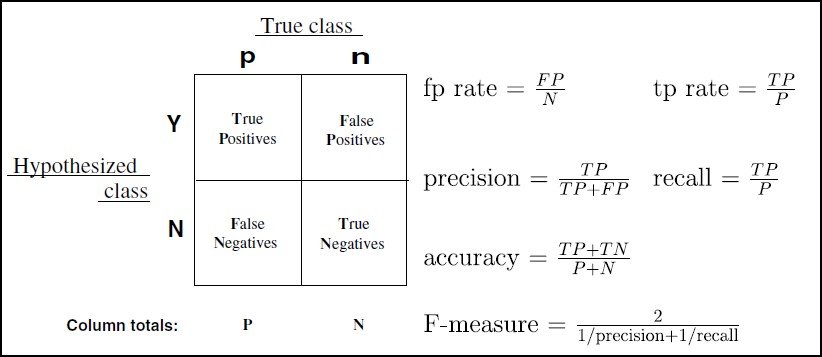


Figure 1. Confusion matrix with common performance metrics. Reprinted from “An introduction to ROC analysis," by T. Fawcett, 2006, Pattern Recognition Letters, 27(8), 862.

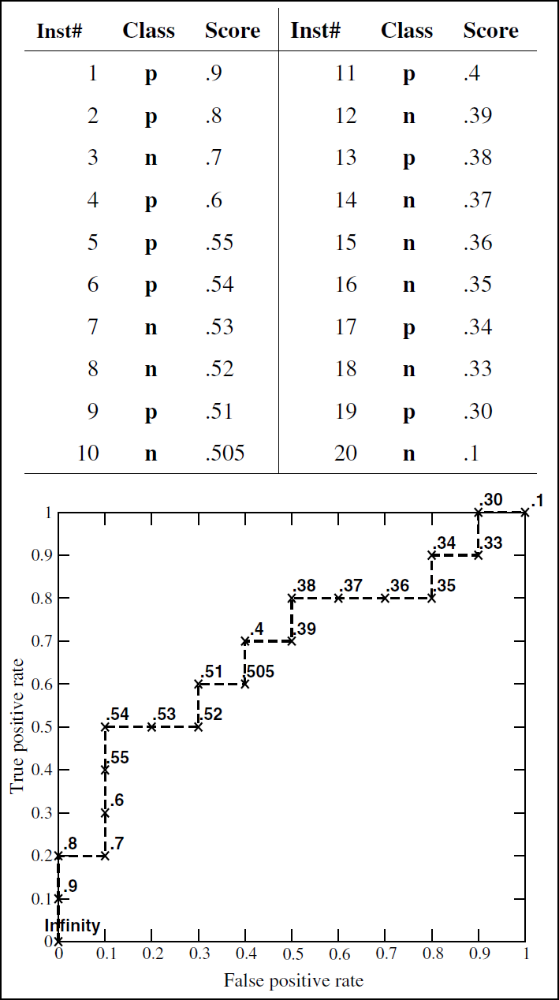


Figure 3. ROC curved created by thresholding a test set. Reprinted from “An introduction to ROC analysis," by T. Fawcett, 2006, Pattern Recognition Letters, 27(8), 862.

More often, we use ROC curves in the context of *ranking*. When ranking, we order a validation set by a value representing the degree to which an instance belongs to a class, known as its *probability* or *score.* Figure 3 shows the score a classifier assigns to each observation in a dataset with 20 records and the corresponding ROC curve with each instant labeled by its probability. ROC curves are invariant to changes in class skew and error costs. If the proportion of negative to positive classes changes in a test set, for instance, the ROC curve remains the same, because its dimensions are based on a single columnar ratio in the confusion matrix. Metrics that use values from both columns, however, including accuracy, precision, the f-measure, and lift, are inherently sensitive to skew (Fawcett, 2006).

*Lift charts* assess how effective our model is at selecting a relatively small sample of records and returning a relatively large sample of the minority class. The input required to generate a lift curve is known as a “scored” data set, a validation data set that includes both probability and actual class columns. We rank these records in descending order according to their propensities. Next, we accumulate the number of actual positive classes and plot this cumulative column on the *y-*axis as a function of the number of records on the *x*-axis. Figure 4 shows a lift curve generated by using the ‘gains’ library in R. For a given amount of observations (the x-axis), the value on the y-axis tells us how much better our model is than chance. The diagonal line in the middle represents the expected number of positive classifications for a given set of records if we select them randomly, shuffling the order of 1’s and 0’s in the actual class column. A lift chart displaying ideal ranking performance would start as a diagonal line with a slope of 1 before leveling off into a horizontal line after accumulating all 1’s. Good classifiers, therefore, give a high lift when acting on just a few records (Shmueli et al., n.d.).

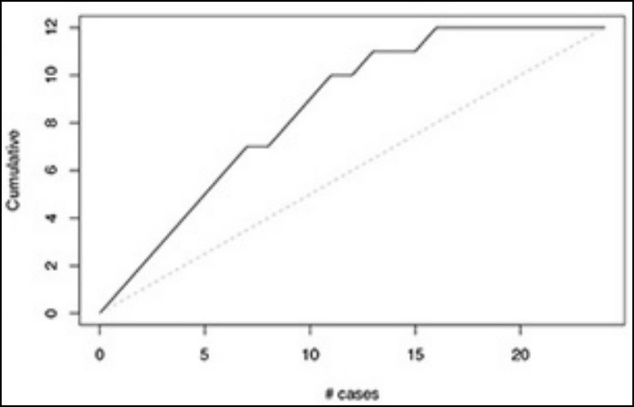


Figure 4. Lift curves generated by the 'gains' library in R. Reprinted from “Data Mining for Business Analytics," by G. Shmueli et al. (n.d.).

In conclusion, the main difference between ROC and lift curves is the metrics they plot (i.e., TPR and FPR vs. cumulative and actual classes). Each uses validation data sets, ranking, and probabilities; the ROC curve is insensitive to skew while the lift curve is not. Lastly, the farther away each curve is from the baseline, the better the predictive capabilities of the model.

References

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., Lichtendahl, K. C., & Jr. (n.d.). Data Mining for Business Analytics. Retrieved from <https://platform.virdocs.com/r/s/0/doc/503437/sp/21743566/mi/74416482?cfi=%2F4%2F2%2F12%2F6%2F4%2F1%3A81>

Sinha, A. P., & May, J. H. (2004). Evaluating and Timing Predictive Data Mining Models Using Receiver Operating Characteristic Curves. Journal of Management Information Systems, 21(3), 249–280. https://doi.org/10.1080/07421222.2004.11045815