



**The Twenty-First International  
Conference on Machine Learning**

**Tutorial on  
"The Many Faces of ROC Analysis in Machine Learning"**

Peter A. Flach  
University of Bristol, UK

Tutorial Notes: PDF | [Part 1](#) | [Part 2](#) | [Part 3](#)

## Topic

Receiver Operating Characteristics (ROC) Analysis originated from signal detection theory, as a model of how well a receiver is able to detect a signal in the presence of noise. Its key feature is the distinction between *hit rate* (or *true positive rate*) and *false alarm rate* (or *false positive rate*) as two separate performance measures. ROC analysis has also widely been used in medical data analysis to study the effect of varying the threshold on the numerical outcome of a diagnostic test. It has been introduced to machine learning relatively recently, in response to classification tasks with varying class distributions or misclassification costs (hereafter referred to as *skew*). ROC analysis is set to cause a paradigm shift in machine learning. Separating performance on classes is almost always a good idea from an analytical perspective. For instance, it can help us to

- understand the behaviour and skew-sensitivity of many machine learning metrics, including rule learning heuristics and decision tree splitting criteria, by plotting their isometrics in ROC space;
- develop new metrics specifically designed to improve the Area Under the ROC Curve (AUC) of a model;
- understand fundamental algorithms such as the separate-and-conquer or sequential covering rule learning algorithm, by tracing its trajectory through a sequence of ROC spaces.

The goal of this tutorial is to develop the ROC perspective in a systematic way, demonstrating the many faces of ROC analysis in machine learning.

# Intended audience

PhD students and machine learning novices will profit from a gentle introduction to ROC analysis for model evaluation and selection and achieve a better understanding of machine learning metrics. Furthermore, they will be motivated to apply the ROC perspective to their own work. More experienced machine learning researchers, who may already be familiar with the more basic material on model evaluation and selection, will benefit from the fresh perspective on machine learning that the tutorial provides, and perhaps be encouraged to tackle some of the open problems in their own research.

After attending this tutorial, participants will be able to

- **model evaluation:** produce ROC plots for categorical and ranking classifiers and calculate their AUC; apply cross-validation in doing so;
- **model selection:** use the ROC convex hull method to select among categorical classifiers; determine the optimal decision threshold for a ranking classifier (calibration);
- **metrics:** analyse a variety of machine learning metrics by means of ROC isometrics; understand fundamental properties such as skew-sensitivity and equivalence between metrics;
- **model construction:** appreciate that one model can be many models from a ROC perspective; use ROC analysis to improve a model's AUC;
- **multi-class ROC:** understand multi-class approximations such as the MAUC metric and calibration of multi-class probability estimators; appreciate the main open problems in extending ROC analysis to multi-class classification.

Only basic knowledge of symbolic classification (decision trees, rule learners) and probabilistic classification (naive Bayes) is expected, at the level of Mitchell or Witten & Frank (one-slide introductions will be provided). It should be noted that most of the techniques covered are not restricted to these classification paradigms.

# Contents and format

The material will be presented in two 90-minute blocks with a break in between. The first half is foundational and covers the basics of ROC analysis from the machine learning literature and related fields. The second half deals with more advanced work including some of my own.

- **Fundamentals (90 minutes)**
  - categorical classification [1]: ROC plots, random selection between models, the ROC convex hull, iso-accuracy lines
  - ranking [2]: ROC curves, the AUC metric, turning rankers into classifiers, calibration, cross-validation
  - interpretation: concavities, majority class performance
  - alternatives: PN plots, cost curves [3], precision-recall diagrams [4], DET curves [5]
- **A broader view (60 minutes)**
  - understanding ML metrics [6,7,8]: isometrics, basic types of linear isometric plots, linear metrics and equivalences between them, non-linear metrics, skew-sensitivity [3,9]
  - model manipulation: obtaining new models without re-training, repairing concavities [10], ordering decision tree branches and rules [11], locally adjusting rankings

- **Multi-class ROC (30 minutes)**

- the general problem, multi-objective optimisation and the Pareto front [12], approximations to Area Under ROC Surface [13], calibrating multi-class probability estimators [14]

The material will mostly be presented using computer-projected PowerPoint slides. In addition, I will make use of the [ROC graph software we developed in Bristol](#).

## Selected references

- [1] F. Provost and T. Fawcett (2001). Robust classification for imprecise environments. *Machine Learning*, 42, 203-231.
- [2] T. Fawcett (2003). [ROC graphs: Notes and practical considerations for data mining researchers](#). Tech report HPL-2003-4. HP Laboratories, Palo Alto, CA, USA.
- [3] C. Drummond and R.C. Holte (2000). Exploiting the cost (in)sensitivity of decision tree splitting criteria. In P. Langley, editor, *Proc. 17th International Conference on Machine Learning (ICML'00)*, pp. 239-246.
- [4] C.J. Van Rijsbergen (1979). *Information retrieval*. London: Butterworths.
- [5] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki (1997). The DET curve in assessment of detection task performance. In *Proc. EuroSpeech*, pp. 1895-1898.
- [6] P.A. Flach (2003). [The geometry of ROC space: understanding machine learning metrics through ROC isometrics](#). In T. Fawcett and N. Mishra, editors, *Proc. 20th International Conference on Machine Learning (ICML'03)*, pp. 194-201. AAAI Press.
- [7] J. Fürnkranz and P.A. Flach (2003). [An analysis of rule evaluation metrics](#). In T. Fawcett and N. Mishra, editors, *Proc. 20th International Conference on Machine Learning (ICML'03)*, pp. 202-209. AAAI Press.
- [8] J. Fürnkranz and P.A. Flach (forthcoming). ROC 'n' rule learning -- towards a better understanding of covering algorithms. *Machine Learning*, accepted for publication.
- [9] R. Vilalta and D. Oblinger (2000). A quantification of distance-bias between evaluation metrics in classification. In P. Langley, editor, *Proc. 17th International Conference on Machine Learning (ICML'00)*, pp. 1087-1094. Morgan Kaufmann.
- [10] P.A. Flach and S. Wu (2003). [Repairing concavities in ROC curves](#). In J.M. Rossiter and T.P. Martin, editors, *Proc. 2003 UK workshop on Computational Intelligence (UKCI'03)*, pp. 38-44. University of Bristol.
- [11] C. Ferri, P.A. Flach, and J. Hernández-Orallo (2002). Learning Decision Trees Using the Area Under the ROC Curve. In C. Sammut and A. Hoffmann, editors, *Proc. 19th International Conference on Machine Learning (ICML'02)*, pp. 139-146. Morgan Kaufmann.

- [12] R.E. Steuer (1986). *Multiple Criteria Optimization: Theory, Computation and Application*. Wiley, New York.
- [13] D.J. Hand and R.J. Till (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems, *Machine Learning*, 45, 171-186.
- [14] N. Lachiche and P.A. Flach (2003). [Improving accuracy and cost of two-class and multi-class probabilistic classifiers using ROC curves](#). In T. Fawcett and N. Mishra, editors, *Proc. 20th International Conference on Machine Learning (ICML'03)*, pp. 416-423. AAAI Press.