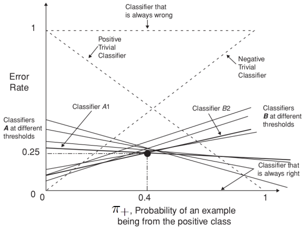
Introduction to Machine Learning

Evaluation: Beyond AUC



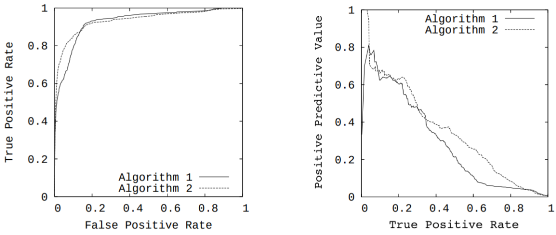
Learning goals

- See the limitations of ROC curves
- Understand the concepts of precision-recall and cost curves

PRECISION-RECALL CURVES

When dealing with highly imbalanced data (i.e., $n_{-} \gg n_{+}$), precision-recall (PR) curves may be more useful than ROC curves:

- Precision = $\rho_{PPV} = \frac{TP}{TP+FP}$, recall = $\rho_{TPR} = \frac{TP}{TP+FN}$ (do not depend on TN).
- Figure (a): ROC curve shows that both algorithms perform well.
- Figure (b): PR curve shows that there is room for improvement (optimum PR curve: top-right corner).
- The PR space reveals that algorithm 2 has an advantage over algorithm 1, which is due to imbalanced classes.



(a) Comparison in ROC space (b) Comparison in PR space

Davis and Goadrich (2006): The Relationship Between Precision-Recall and ROC Curves (URL).

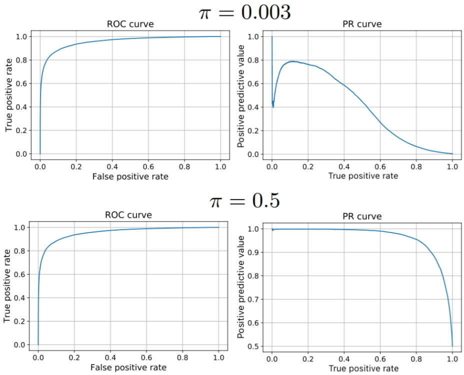
PRECISION-RECALL CURVES

- The issue with ROC is the ρ_{FPR} in case of $n_- \gg n_+$, which is often very small as the TN are usually high.
  $\Rightarrow$ A large change in FP yields a small change in the ρ_{FPR}.
- Precision: compares FP to TP (not TN) and does not suffer from this issue. It captures the effect of the large n_-.

	Actual Positive	Actual Negative	
Predicted Positive	**True positive (TP)**	**False positive (FP, Type I error)**	Precision (P) = Positive predictive value $= \frac{\#TP}{\#TP + \#FP}$
Predicted Negative	**False negative (FN, Type II error)**	**True negative (TN)**	Negative predictive value $= \frac{\#TN}{\#FN + \#TN}$
	Sensitivity = Recall (R) = True positive rate (tpr) $= \frac{\#TP}{\#TP + \#FN}$	Specificity = True negative rate (tnr) $= \frac{\#TN}{\#FP + \#TN}$	Accuracy $= \frac{\#TP + \#TN}{n}$ Error rate $= \frac{\#FN + \#FP}{n}$

PRECISION-RECALL CURVES

- Figures in the top row concern imbalanced classes with $\pi = 0.003$, those in the bottom row balanced ones with $\pi = 0.5$.
- Task and learners remain the same, only class distribution has changed.
- ROC curves (left) are similar, while the PR curve (right) changes dramatically from imbalanced to balanced classes.



Wissam Siblini et. al. (2004): Master your Metrics with Calibration (URL).

PRECISION-RECALL CURVES

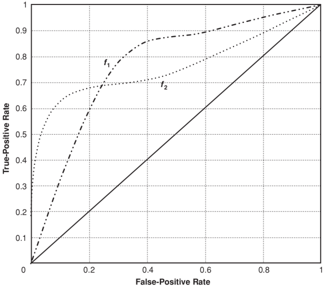
Conclusion:

- ROC and PR curves for given algorithms contain the same points.
- Analogous to the convex hull in ROC space, there is a convex hull in PR space (same points omitted as in convex hull in ROC space).
- An algorithm that optimizes the AUROC is not guaranteed to optimize the AUPRC.
  - Use ROC curves when the numbers of samples in both classes are roughly equal.
  - Use PR curves when there is a moderate to large class imbalance and predicting positive instances is more relevant.

COST CURVES

- **Cost curves** directly plot the relative costs / misclassification error to determine the best classifier (ROC isometrics allow this only indirectly).
- Cost curves incorporate similar information as ROC curves but are easier, especially in case of different misclassification costs or class distributions.

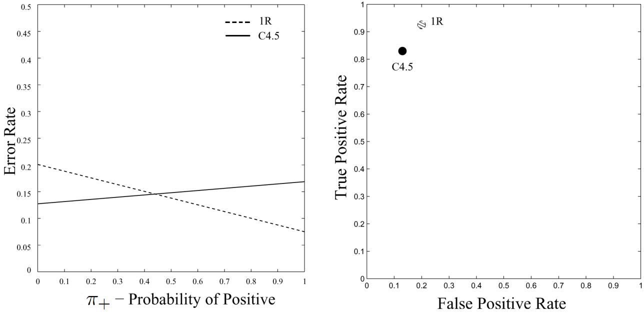
Example: Classifier f_2 dominates f_1 until both ROC curves cross. Then, f_1 dominates f_2.
BUT: It is hard to tell for which threshold, costs, or class distributions f_2 works better than f_1.
$\Rightarrow$ Cost curves provide this kind of information.



Nathalie Japkowicz (2004): Evaluating Learning Algorithms : A Classification Perspective. (p. 125)

COST CURVES

- Simplifying assumption: equal misclassification costs, i.e., $cost_{FN} = cost_{FP}$.
- Misclassification cost (or misclassification error in the case of $cost_{FN} = cost_{FP}$) is plotted as a function of the proportion of positive instances, π_+.
- Cost curves are point–line duals of ROC curves, i.e., a single classifier is represented by a point in the ROC space and by a line in the cost space.

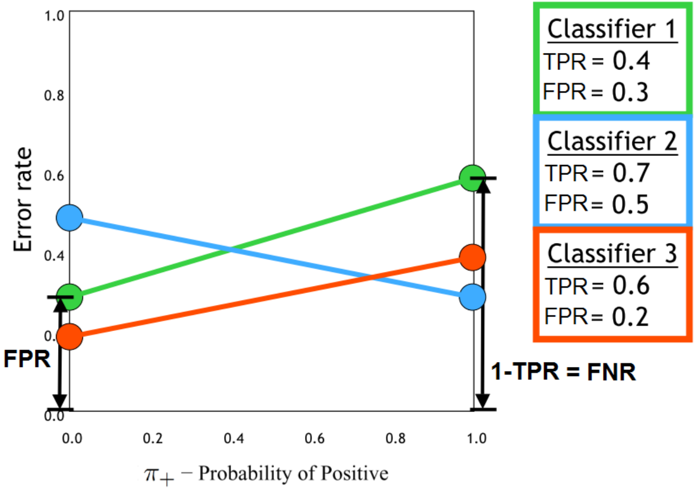


Chris Drummond and Robert C. Holte (2006): Cost curves: An improved method for visualizing classifier performance. Machine Learning, 65, 95-130 (URL).

COST CURVES

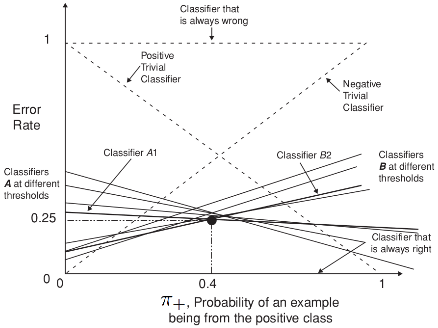
Functional form of the cost curve of a classifier:

$$\textit{error} = (\rho_{FNR} - \rho_{FPR}) \cdot \pi_+ + \rho_{FPR}.$$



COST CURVES

- Horizontal dashed line: worst classifier, i.e. 100% error rate for all π_+; x-axis as perfect classifier.
- Dashed diagonal lines: trivial classifiers, i.e., ascending diagonal always predicts negative instances and vice versa.
- Descending/ascending bold lines: two families of classifiers *A* and *B* (represented by points in their respective ROC curves).



Credit: Nathalie Japkowicz

ROC CURVES VS. COST CURVES

- In some cases, cost curves provide more practical relevant information than ROC curves.
- ROC curves can tell us that, sometimes, classifier A is superior to B, but we cannot really tell in which settings.
- Here, we assumed that the classes had similar classification costs.
- However, there is an extension to the cost curve concept which allows for different costs in each class ($\rightarrow$ simple modification where the identity of the axes is changed).