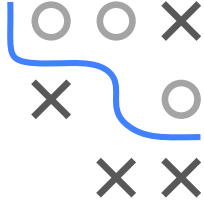
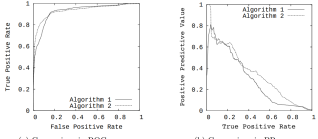
Introduction to Machine Learning



Evaluation
Precision-Recall Curves



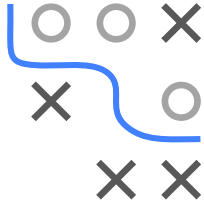
(a) Comparison in ROC space

(b) Comparison in PR space

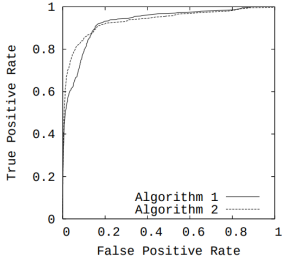
Learning goals

- Understand PR curves
- Same as PPV-TPR curve
- Compare to standard TPR-FPR ROC curve

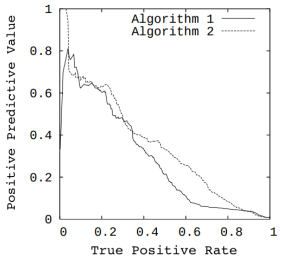
PRECISION-RECALL CURVES



- Slightly changed ROC plot
- Simply plot precision and recall, instead of TPR-FPR
- Precision = $\rho_{PPV} = \frac{TP}{TP+FP}$, recall = $\rho_{TPR} = \frac{TP}{TP+FN}$
- Might call them TPR-PPV curve
- NB: Both metrics don't depend on TNs



(a) Comparison in ROC space

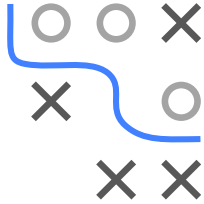


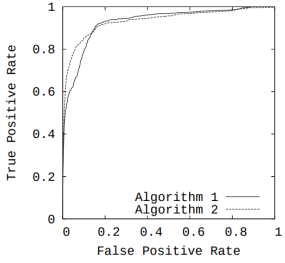
(b) Comparison in PR space

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

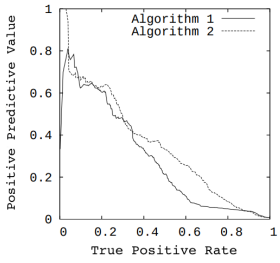
PRECISION-RECALL CURVES

- Might be better for highly imbal data ($n_- \gg n_+$) than TPR-FPR
- Figure (a): ROC; both learners seem to perform well
- Figure (b): PR; visible room for improvement (top-right=best)
- PR reveals better that algo 2 has advantage over 1





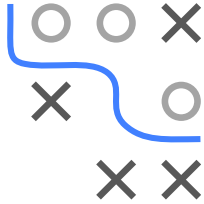
(a) Comparison in ROC space



(b) Comparison in PR space

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

IMBALANCED DATA



- Assume imbalanced classes with $n_- \gg n_+$
- If neg class large, typically less interested in high TNR = low FPR, but more in PPV
- Large (abs) change in FP yields small change in FPR
- PPV likely more informative

FP=10:

	True +1	True -1
Pred. Pos	100	10
Pred. Neg	10	9990
Total	110	10000

$\text{TPR} = 10/11$
$\text{FPR} = 0.001$
$\text{PPV} = 10/11$

FP=100:

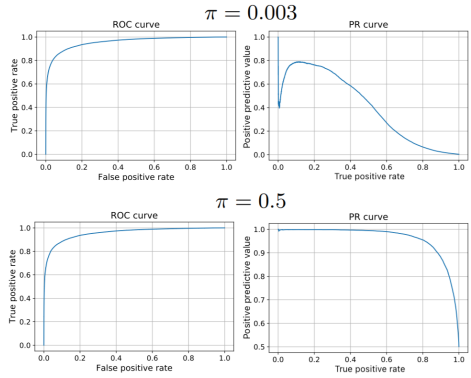
	True +1	True -1
Pred. +1	100	100
Pred. -1	10	9900
Total	110	10000

$\text{TPR} = 10/11$
$\text{FPR} = 0.01$
$\text{PPV} = 1/2$

RHS: Given test says +1, it's now a coin flip that this is correct.

IMBALANCED DATA / 2

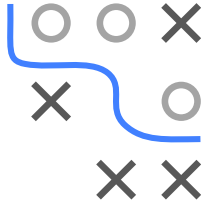
- Top row: Imbal classes with $\pi = 0.003$
- Bottom: balanced with $\pi = 0.5$
- ROC curves (LHS) are similar
- PR curve (RHS) changes strongly from imbal to bal classes

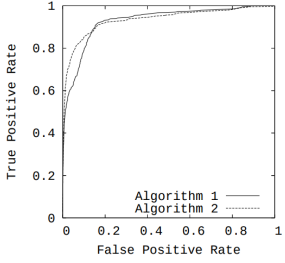


Wissam Siblini et. al. (2004): Master your Metrics with Calibration (URL).

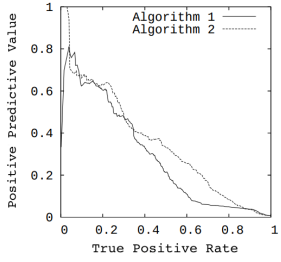
CONCLUSIONS

- Curve fully dominates in ROC space iff dominates in PR-space
- In imbalanced situations rather use PR than standard TPR-FPR
- If comparing few models on a single task, probably plot both. Then observe and think.
- For tuning: can also use PR-AUC (or partial versions)





(a) Comparison in ROC space



(b) Comparison in PR space

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