Performance Measures

Sören Sonnenburg

FIRST

Fraunhofer Institut Rechnerarchitektur und Softwaretechnik

Fraunhofer FIRST.IDA, Kekuléstr. 7, 12489 Berlin, Germany

Roadmap:

- Contingency Table
- Scores from the Contingency Table
- Curves from the Contingency Table
- Discussion

Contingency Table

Contingency Table / Confusion Matrix

Given

- sample $\boldsymbol{x}_1 \dots \boldsymbol{x}_N \subset \mathcal{X}$
- 2-class classifier $f(x) \mapsto \{-1,+1\}$, outputs $f(\boldsymbol{x}_1) \dots f(\boldsymbol{x}_n)$ for $\boldsymbol{x}_1 \dots \boldsymbol{x}_N$
- true labeling $(y_1 \dots y_N) \in \{-1,+1\}^N$

$\Rightarrow$ partitions the $N-$sample into

outputs\ labeling	y=+1	y=-1	Σ
f(x)=+1	TP	FP	O^+
f(x)=-1	FN	TN	O^-
Σ	N^+	N^-	N

Scores I

o\ l	y=+1	y=-1	Σ
f(x)=+1	TP	FP	O^+
f(x)=-1	FN	TN	O^-
Σ	N^+	N^-	N

- $acc = \frac{TP+TN}{N}$ - accurracy
- $err = \frac{FP+FN}{N}$ - error

Scores II

o\ l	y=+1	y=-1	Σ
f(x)=+1	TP	FP	O^+
f(x)=-1	FN	TN	O^-
Σ	N^+	N^-	N

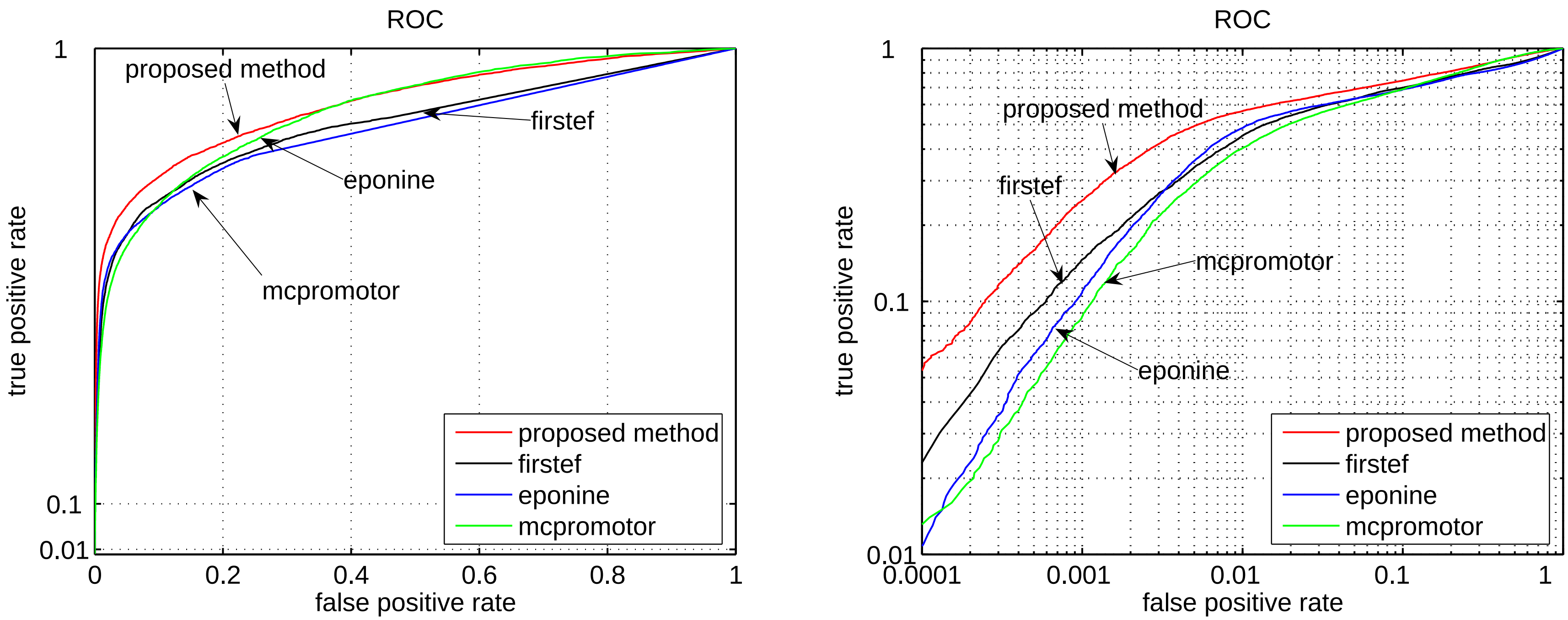
- $tpr = TP/N^+ = \frac{TP}{TP+FN}$ - sensitivity/recall
- $tnr = TN/N^- = \frac{TN}{TN+FP}$ - specificity
- $fnr = FN/N^+ = \frac{FN}{FN+TP}$ - 1-sensitivity
- $fpr = FP/N^- = \frac{FP}{FP+TN}$ - 1-specificity
- $ppv = TP/O^+ = \frac{TP}{TP+FP}$ - positive predictive value / precision
- $fdr = FP/O^+ = \frac{FP}{FP+TP}$ - false discovery rate
- $npv = TN/O^- = \frac{TN}{TN+FN}$ - negative predictive value
- $fdr_- = FN/O^- = \frac{FN}{FN+TN}$ - negative false discovery rate

Scores III

o\ l	y=+1	y=-1	Σ
f(x)=+1	TP	FP	O^+
f(x)=-1	FN	TN	O^-
Σ	N^+	N^-	N

- $ber = \frac{1}{2}\left(\frac{FN}{FN+TP} + \frac{FP}{FP+TN}\right)$ - balanced error
- $WRacc = \frac{TP}{TP+FN} - \frac{FP}{FP+TN}$ - weighted relative accuracy
- $f1 = \frac{2*TP}{2*TP+FP+FN}$ - F1 score (harmonic mean between precision/recall)
- $cc = \frac{TP^2 - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$ - cross correlation

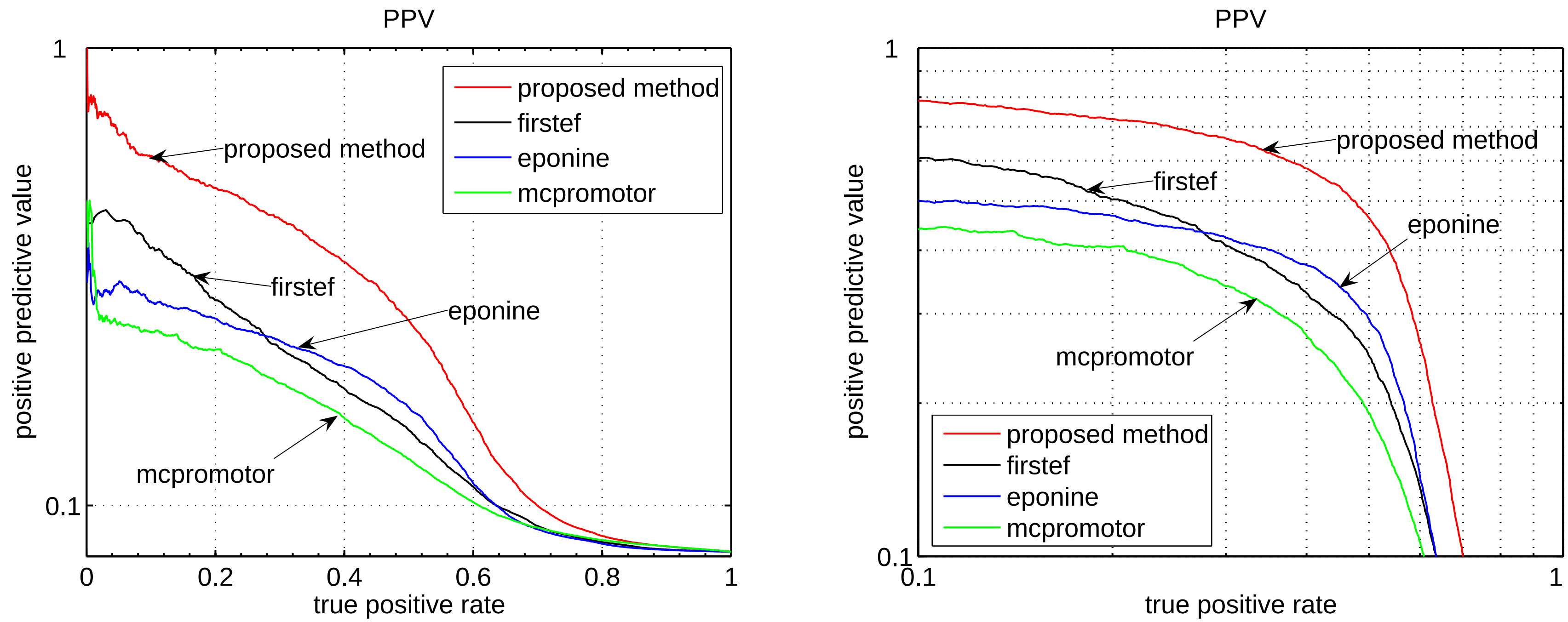
Receiver Operator Characteristic Curve



- use tpr/fpr
- obtained by varying bias
- independent of class skew
- monotonely ascending

CURVES

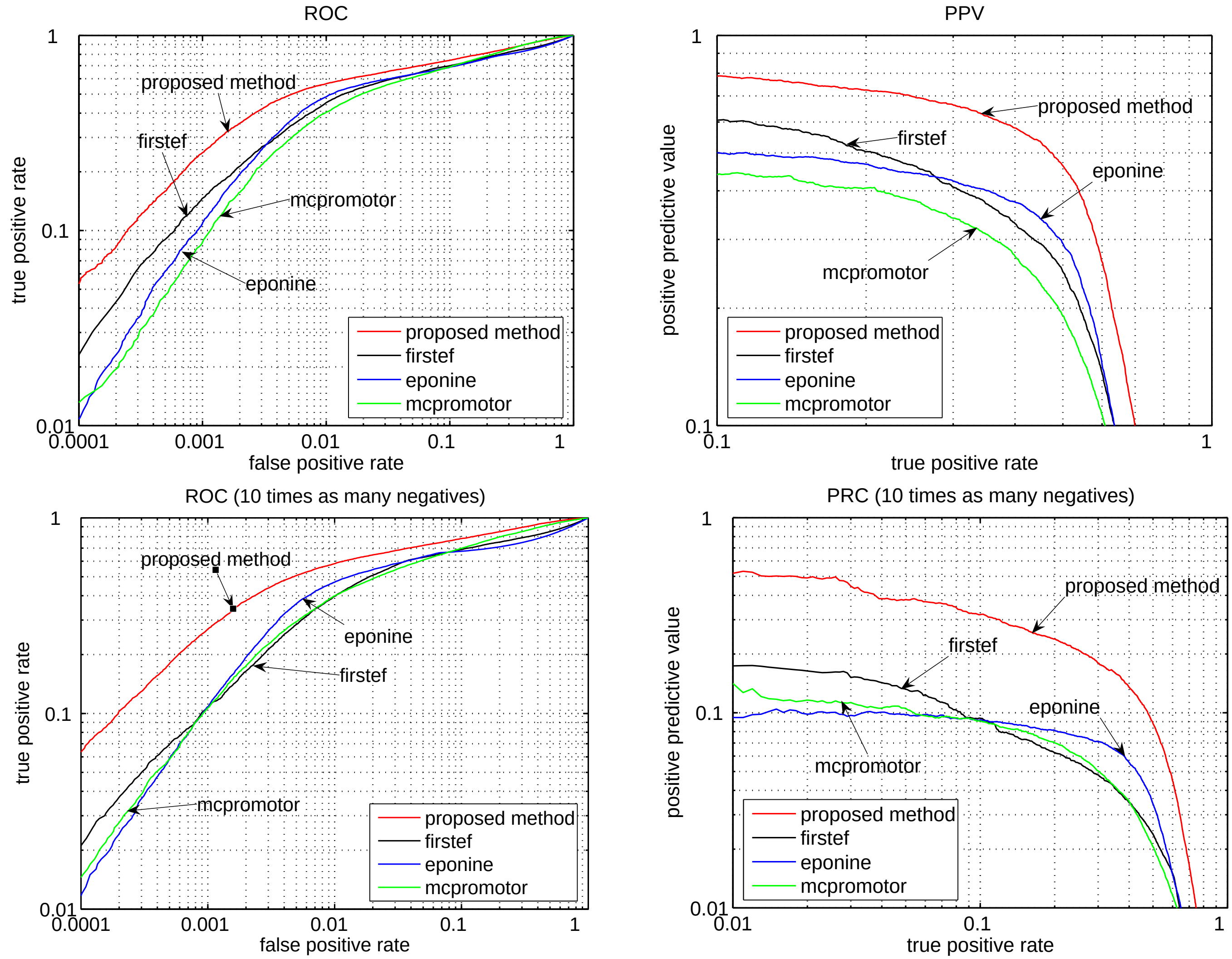
Precision Recall Curve



- use ppv/tpr
- obtained by varying bias
- depends on class skew
- not monotone

CURVES

ROC vs. PRC



Area under the Curves

auROC remember: $tpr = TP/N^+ = \frac{TP}{TP+FN}$

and $fpr = FP/N^- = \frac{FP}{FP+TN}$

Properties

- considering output as ranking, it is the number of swappings such that output of positive examples $\geq$ output of negative examples divided by $N^+ \cdot N^-$
- equivalent to wilcoxon rank test
- $Gini = 2 \cdot auROC - 1$
- independant of bias
- independant of class skew

Area under the Curves

auPRC remember: $ppv = TP/O^{+} = \frac{TP}{TP+FP}$

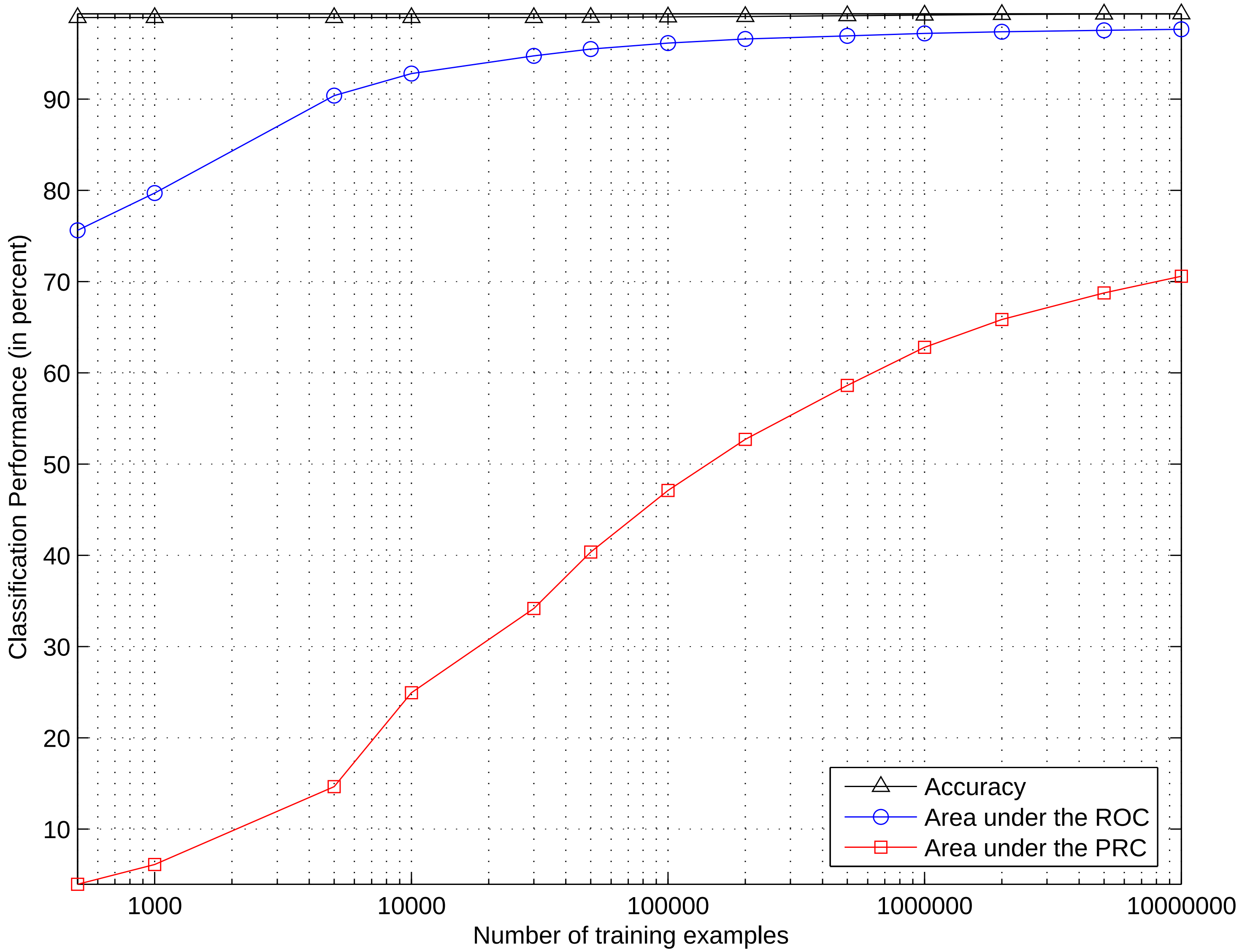
and $tpr = TP/N^{+} = \frac{TP}{TP+FN}$

Properties

- considering output as ranking, . . . ???
- meaning ???
- independant of bias
- *dependant* on class skew

Comparison

accuracy, auROC, auPRC



Extensions

- ROC_N Score - area under the ROC curve up to the first N false-positives
- other ???

Summary

When use what:

- balanced data $\Rightarrow$ accuracy/auROC/. . .
- unbalanced date $\Rightarrow$ BER/F1/auPPV/CC
- other cases / other measures ???

auROC/auPRC optimized training

- current classifiers minimize trainerror (+ some complexity)
- want to optimize auROC/auPRC/... directly
- first approaches (also implemented in shogun) create pairs of examples $(x_I, x_J), I = \{i|y_i = +1\}, J = \{i|y_i = +1\}$, $\Rightarrow N^+\dot{N}^-$,e.g. $O(N^2)$ examples in standard SVM learning $\Rightarrow$ unusuably slow
- recently TJ (ICML 2005) SVM$^{\Delta}_{multi}$:

$$\min_{\boldsymbol{w},\xi\geq 0} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C\xi$$

s.t. $\forall \boldsymbol{Y} \in \mathcal{Y}\backslash\boldsymbol{Y} : \boldsymbol{w}^T[\Psi(\boldsymbol{X},\boldsymbol{Y}) - \Psi(\boldsymbol{X},\boldsymbol{Y}')] \geq \Delta(\boldsymbol{Y}',\boldsymbol{Y}) - \xi$

with $\Psi(\boldsymbol{X},\boldsymbol{Y}) = \sum_{i=1}^{n} y_i \boldsymbol{x}_i$

AIM

- has been done for auROC, F1, BER and is doable for any other measure based on scores in contingency table
- AIM: Do this for auPRC/ROC_n !

$$\min_{\boldsymbol{w},\xi\geq 0} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C\xi$$

$$\text{s.t.} \quad \forall \boldsymbol{Y} \in \mathcal{Y}\backslash\boldsymbol{Y} : \boldsymbol{w}^T[\Psi(\boldsymbol{X},\boldsymbol{Y}) - \Psi(\boldsymbol{X},\boldsymbol{Y}')] \geq \Delta(\boldsymbol{Y}',\boldsymbol{Y}) - \xi$$

$$\text{with} \quad \Psi(\boldsymbol{X},\boldsymbol{Y}) = \sum_{i=1}^{n} y_i \boldsymbol{x}_i$$

Anyone interested ?

Outlook

- muliclass
- regression
- ...
- ...