SVMpAUC-tight: A New Support Vector Method for Optimizing Partial AUC Based on a *Tight Convex Upper Bound*

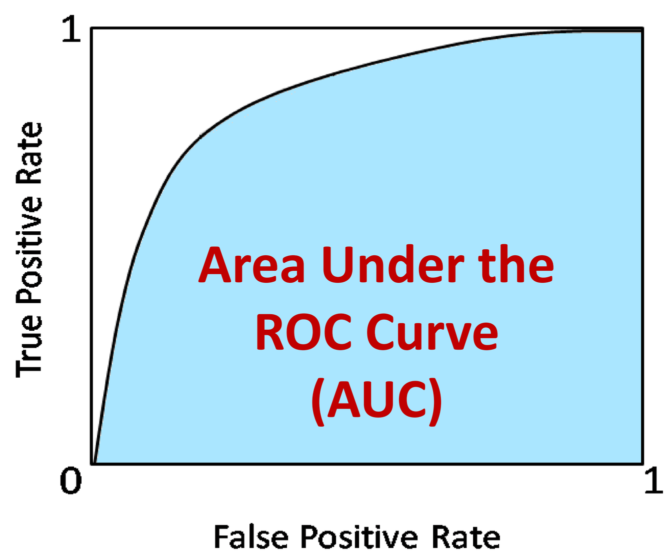
Harikrishna Narasimhan and Shivani Agarwal

Department of Computer Science and Automation, Indian Institute of Science, Bangalore

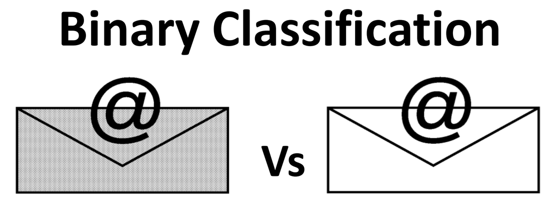
Abstract

The area under the ROC curve (AUC) is a widely used performance measure in machine learning and data mining. However, in several applications, performance is measured not in terms of the full AUC, but instead in terms of the *partial* AUC between two specified false positive rates. In recent work, we proposed a structural SVM based approach for optimizing this performance measure (Narasimhan and Agarwal, 2013). In this paper, we develop a new support vector method, SVMpAUC-tight, that optimizes a tighter convex upper bound on the partial AUC loss, which leads to *improved accuracy* and *reduced computational complexity*. As with our previous method, the resulting optimization problem is solved using a cutting plane method. We demonstrate the effectiveness of the new method on a variety of bioinformatics tasks. In addition, using a projected subgradient solver, we develop extensions of our method to learn sparse and group sparse models.

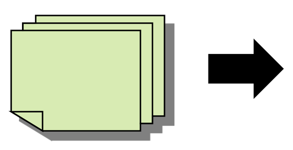
Receiver Operating Characteristic Curve



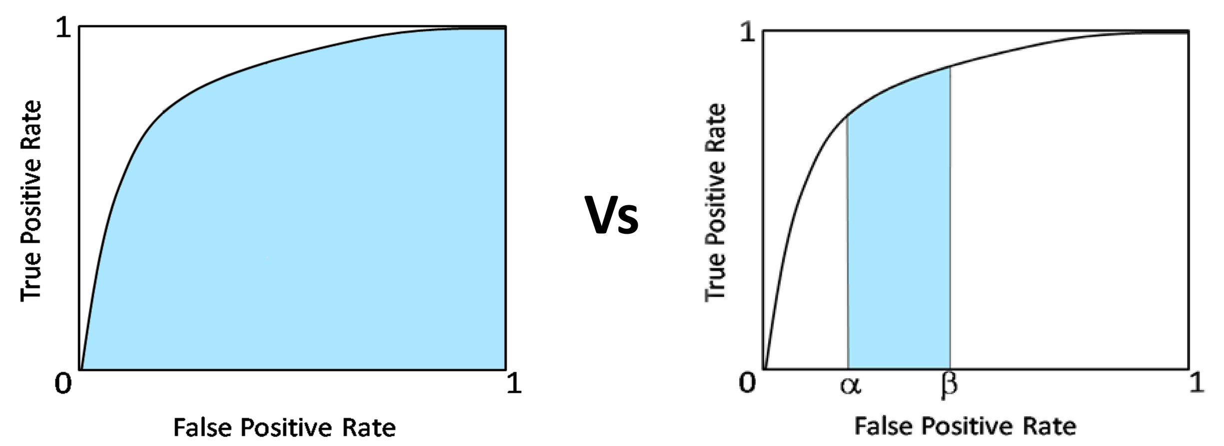
Binary Classification



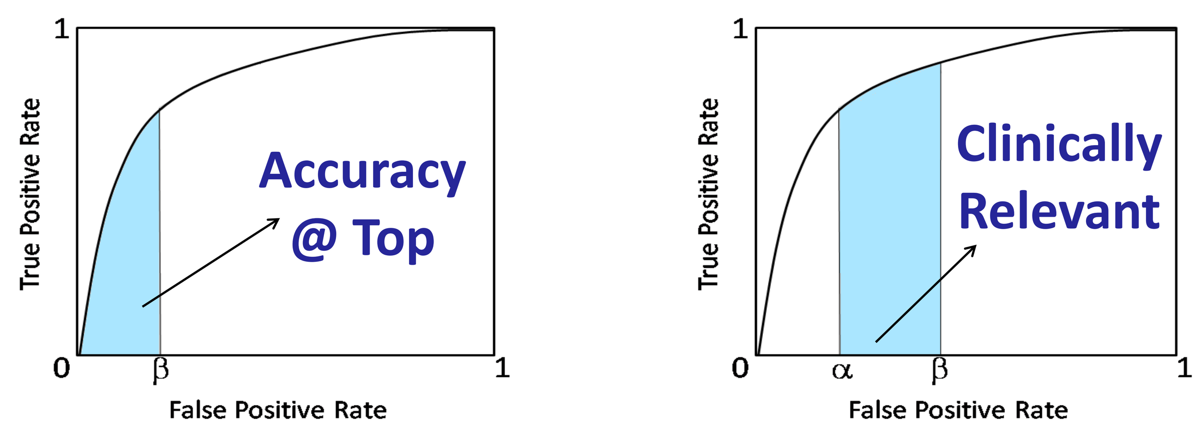
Bipartite Ranking



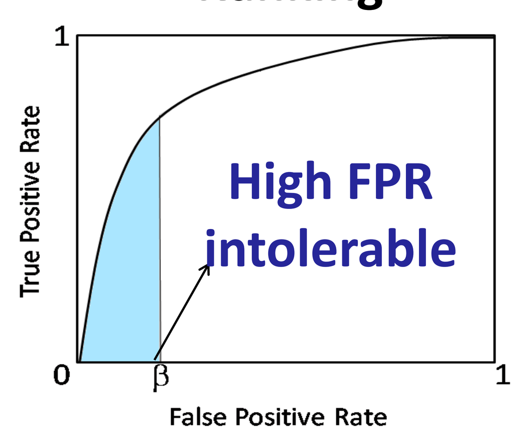
Full AUC Vs. Partial AUC



Applications



Ranking **Medical Diagnosis**



Biometric Screening

KDD Cup 2008

(Early Breast Cancer Detection)

Evaluation metric:

Scaled version of Partial AUC in [α, β]

Problem Setup

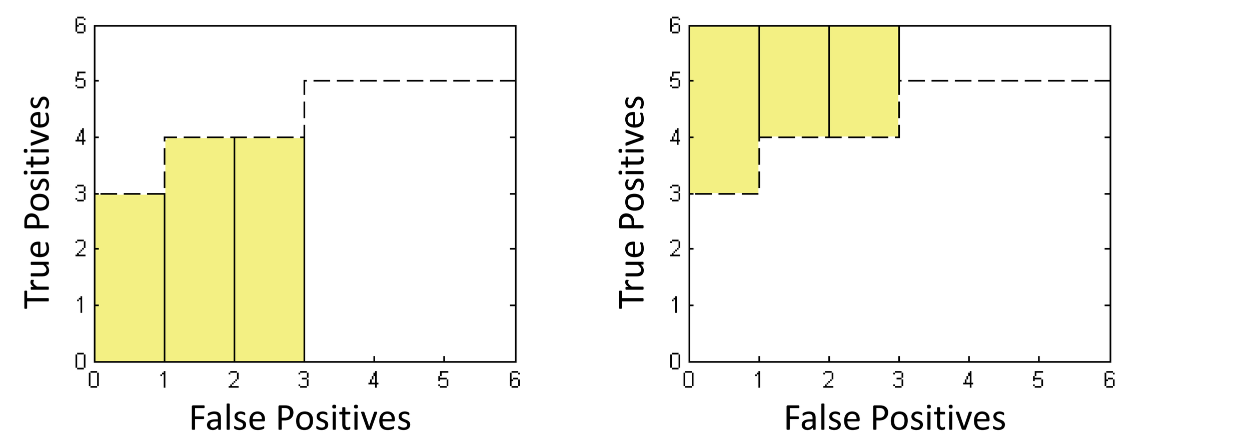
Positive Instances x_1^+ x_2^+ x_3^+ .... x_m^+

Negative Instances x_1^- x_2^- x_3^- .... x_n^-

GOAL? Learn a scoring function $f : X \rightarrow \mathbb{R}$

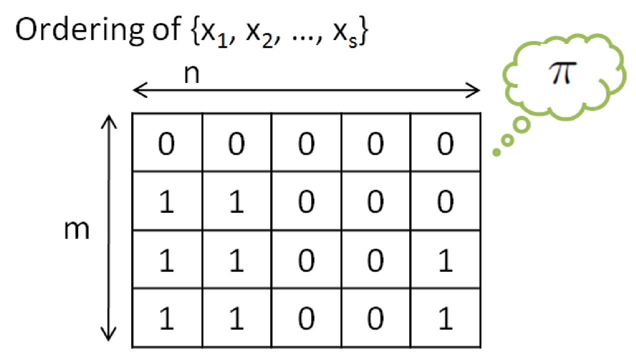
Optimize: (1 – pAUC) for f

Partial AUC



$$(1-\text{pAUC}) \propto \sum_{\text{negatives } j \text{ in FPR range } [\alpha,\beta]}^{n} \left(\text{no. of positives below } j^{\text{th}} \text{ negative}\right)$$

SVMpAUC-struct (ICML 2013)



(1 – pAUC) for f

≤

$$\max_{\text{ordering matrices } \pi \text{ of size } m\times n} \left(\underset{\text{for } \pi}{(1-\text{pAUC})} + \underset{\text{on all pairs}}{\text{term capturing agreement between } \pi \text{ and } f} \right)$$

+ Regularizer

Pair-wise Hinge Loss

$$(1-\text{pAUC}) \propto \sum_{\substack{\text{negatives } j \text{ in}\\ \text{FPR range } [\alpha,\beta]}}^{n} \sum_{i=1}^{m} \mathbf{1}\left(f(x_i^+) - f(x_j^-) \leq 0\right)$$

$$\leq \sum_{\substack{\text{negatives } j \text{ in}\\ \text{FPR range } [\alpha,\beta]}}^{n} \sum_{i=1}^{m} \text{hinge-loss}\left(f(x_i^+) - f(x_j^-)\right)$$

Upper Bound for FPR range [0, β]

Subset of pairs of positive-negative examples

$$\max_{\text{ordering matrices } \pi \text{ of size } m\times n} \left(\underset{\text{for } \pi}{(1-\text{pAUC})} + \underset{\text{on \underline{all pairs}}}{\text{term capturing agreement between } \pi \text{ and } f} \right)$$

=

pair-wise hinge loss + extra term

Upper Bound for FPR range [α, β)

approx. pair-wise hinge loss + extra term

≤

$$\max_{\text{ordering matrices } \pi \text{ of size } m\times n} \left(\underset{\text{for } \pi}{(1-\text{pAUC})} + \underset{\text{on all pairs}}{\text{term capturing agreement between } \pi \text{ and } f} \right)$$

≤

pair-wise hinge loss + extra term

SVMpAUC-tight: Improved Formulation

$$(1-\text{pAUC}) \propto \max_{\substack{\text{subsets } S \text{ of negatives}\\ \text{of size } j_\beta}} \left(1-\text{pAUC restricted to negatives in } S\right)$$

SVMpAUC-struct objective restricted to S

$$\max_{\substack{\text{subsets } S \text{ of negatives}\\ \text{of size } j_\beta}} \left[\max_{\substack{\text{ordering matrices } \pi\\ \text{of size } m\times j_\beta}} \left(\underset{\text{for } \pi}{\text{truncated } (1-\text{pAUC})} + \underset{\text{on pairs corresponding to } S}{\text{term capturing agreement between } \pi \text{ and } f} \right) \right]$$

Tighter upper bound: extra term absent

SVMpAUC-tight: Optimization Problem

$$\min_{w,\xi\geq 0} \frac{1}{2}\|w\|_2^2 + C\xi$$

$$\text{s.t. } \forall z \in \mathcal{Z}_\beta,\ \pi \in \Pi_{m,j_\beta}:$$

$$\boxed{w^\top\left(\phi_z(S,\pi^*) - \phi_z(S,\pi)\right) \geq \Delta_\beta(\pi^*,\pi) - \xi}$$

Exponential number of constraints!

Cutting Plane Solver

Repeat:

1. Solve OP for a subset of constraints.
2. Add the most violated constraint.

Better Runtime Guarantee:

Max. number of iterations

Time taken per iteration

SVMpAUC-tight: Sparse Extensions

Primal Formulation

$$\min_w \left[\max_{z\in\mathcal{Z}_\beta,\ \pi\in\Pi_{m,j_\beta}} \Delta_\beta(\pi^*,\pi) - w^\top\left(\phi_z(S,\pi^*) - \phi_z(S,\pi)\right) \right]$$

s.t.

$$\boxed{\|w\|_2 \leq \lambda}$$

Projected Subgradient

Repeat:

1. Compute subgradient and perform update.
2. Project on to the constraint set.

Sparsity-inducing Regularizations:

LASSO

Group LASSO

Elastic-Net

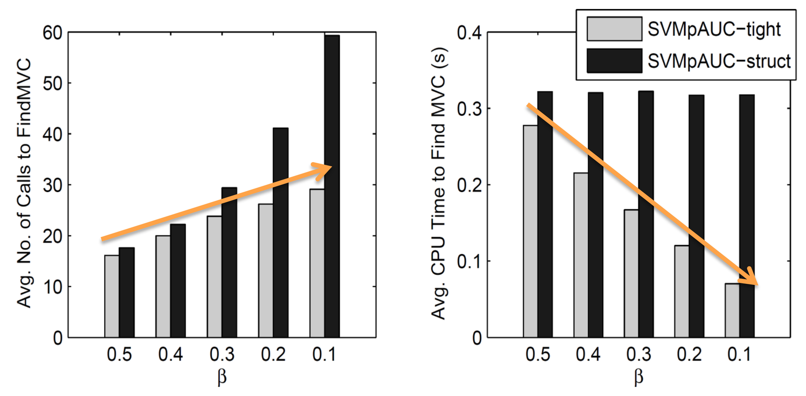
Experimental Results

Partial AUC in [0, 0.1]

	PPI	Chem-informatics	KDD Cup 2001	Leukemia	Ovarian Cancer
SVMpAUC-tight	**52.95**	**65.30**	69.91	**30.44**	91.84
SVMpAUC-struct	51.96	65.28	**70.12**	24.64	91.84
SVMAUC	39.72	62.78	62.23	28.83	**92.17**

Partial AUC in [0.2s, 0.3s]

	KDD Cup 2008
SVMpAUC-tight	**53.43**
SVMpAUC-struct	51.89
SVMAUC	50.66



Cutting Plane Vs Projected Subgradient

CP faster on high dim. data with L2 regularization

PSG faster with L1 regularization

Sparse Extensions

Elastic-net regularization yields sparse models with pAUC comparable to L2

References

1. T. Joachims. A support vector method for multivariate performance measures. *ICML*, 2005.
2. T. Joachims, T. Finley, and C.-N. J. Yu. Cutting-plane training of structural SVMs. *Machine Learning*, 77(1):27–59, 2009.
3. H. Narasimhan and S. Agarwal. A structural SVM based approach for optimizing partial AUC. *ICML*, 2013.