

# A Structural SVM Based Approach for Optimizing the Partial AUC

# 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. 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 this paper, we develop a structural SVM framework for directly optimizing the partial AUC between any two false positive rates. Our approach makes use of a cutting plane solver along the lines of the approach for optimizing the full AUC developed by Joachims (2005). One of our key contributions is an efficient algorithm for solving the combinatorial optimization problem needed to find the most violated constraint in the cutting plane solver. We demonstrate the effectiveness of our approach on a variety of real-world tasks.



#### **Structural SVM Based Formulation**

## **Receiver Operating Characteristic Curve**















#### **Experiments**

#### **Drug Discovery** 50 active compounds / 2092 inactive compounds

	pAUC(0, 0.1)
$\mathrm{SVM}_{\mathrm{pAUC}}[0,0.1]$	65.25
$SVM_{AUC}$	62.64 *
ASVM[0,0.1]	63.80
pAUCBoost[0,0.1]	43.89 *
Greedy-Heuristic[0,0.1]	8.33 *

#### **KDD Cup 08 - Breast Cancer Detection** ~600 malignant ROIs / ~10<sup>5</sup> benign ROIs

	PAUC(0.2s, 0.3s)
$SVM_{pAUC}[0.2s, 0.3s]$	51.44
$SVM_{AUC}$	50.50
pAUCBoost[0.2s, 0.3s]	48.06 *
Greedy-Heuristic $[0.2s, 0.3s]$	46.99 *



### Scaled version of Partial AUC in $[\alpha, \beta]$

**Biometric Screening** 

# **Partial AUC Optimization**

- Existing techniques: ASVM, pAUCBoost, Several heuristic methods.
- Many existing approaches heuristic in nature or handle special cases of the problem.

#### **Our Contribution**

A new support vector method for optimizing the general partial AUC.

Based on Joachims' Structural SVM approach for full AUC optimization, but leads to a trickier inner combinatorial optimization problem.

#### **Protein-Protein Interaction Prediction**

~3x10<sup>3</sup> interacting pairs / ~2x10<sup>5</sup> non-interacting pairs

	pAUC(0, 0.1)
$SVM_{pAUC}[0,0.1]$	51.79
$SVM_{AUC}$	39.72 *
ASVM[0,0.1]	44.51 *
pAUCBoost[0,0.1]	48.65 *
Greedy-Heuristic[0,0.1]	47.33 *

#### **Run Time Analysis**



# References

- 1. T. Joachims. A Support Vector Method for Multivariate Performance Measures. In ICML, 2005.
- 2. T. Joachims. Training linear SVMs in linear time. In KDD, 2006.
- 3. S.-H. Wu, K.-P. Lin, C.-M. Chen, and M.-S. Chen. Asymmetric support vector machines: low false-positive learning under the user tolerance. In KDD, 2008.
- 4. O. Komori and S. Eguchi. A boosting method for maximizing the partial area under the ROC curve. BMC *Bioinformatics*, 11:314, 2010.