
A Structural SVM Based Approach for Optimizing Partial AUC

Supplementary Material

Harikrishna Narasimhan
Shivani Agarwal

HARIKRISHNA@CSA.IISC.ERNET.IN
SHIVANI@CSA.IISC.ERNET.IN

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

This document contains supplementary material for the paper “A Structural SVM Based Approach for Optimizing Partial AUC”, in *Proceedings of the 30th International Conference on Machine Learning*, 2013.

A. Cutting Plane Method for Optimizing Partial AUC

Algorithm 2 contains an outline of the cutting plane method (for optimizing partial AUC) described in Section 3 of the paper.

Algorithm 2 Cutting Plane Method for SVM_{pAUC}

- 1: **Inputs:** $S = (S_+, S_-)$, α , β , C , ϵ
 - 2: **Initialize:** $C = \emptyset$
 - 3: $H(S, \pi; w) \equiv \Delta_{\text{pAUC}(\alpha, \beta)}(\pi^*, \pi) - \frac{1}{mn(\beta - \alpha)} \sum_{i,j} \pi_{ij} w^\top (x_i^+ - x_j^-)$
 - 4: **Repeat**
 - 5: $(w, \xi) = \underset{w, \xi \geq 0}{\operatorname{argmin}} \frac{1}{2} \|w\|^2 + C\xi$
s.t. $\forall \pi \in C : \xi \geq H(S, \pi; w)$
 - 6: $\bar{\pi} = \underset{\pi \in \Pi_{m,n}}{\operatorname{argmax}} H(S, \pi; w)$
(compute the most violated constraint)
 - 7: $C = C \cup \{\bar{\pi}\}$
 - 8: **Until** $(H(S, \bar{\pi}; w) \leq \xi + \epsilon)$
 - 9: **Output:** w
-

B. Proof of Theorem 1

This section contains the proof for Theorem 1 in Section 4 of the paper.

Theorem 1. The solution $\bar{\pi}$ to OP2 lies in $\Pi_{m,n}^w$.

Proof. Suppose $\bar{\pi} \notin \Pi_{m,n}^w$. Then $\exists i, j_1 < j_2$ such that $\bar{\pi}_{i,(j_1)_w} < \bar{\pi}_{i,(j_2)_w}$, i.e. such that $\bar{\pi}_{i,(j_1)_w} = 0$ and $\bar{\pi}_{i,(j_2)_w} = 1$. This means that $\bar{\pi}$ ranks x_i^+ above $x_{(j_1)_w}^-$ but below $x_{(j_2)_w}^-$. Now let us construct from $\bar{\pi}$ an ordering $\bar{\pi}'$ in which the instances $x_{(j_1)_w}^-$ and $x_{(j_2)_w}^-$ are swapped, i.e. for all i' with $\bar{\pi}_{i',(j_1)_w} = 0$ and $\bar{\pi}_{i',(j_2)_w} = 1$, we set $\bar{\pi}'_{i',(j_1)_w} = 1$ and $\bar{\pi}'_{i',(j_2)_w} = 0$.

Then it can be seen that while the loss term in the objective in OP2 is the same for $\bar{\pi}'$ as for $\bar{\pi}$, the second term increases, yielding a higher objective value, $Q_w(\bar{\pi}') > Q_w(\bar{\pi})$. This contradicts the fact that $\bar{\pi}$ is a maximizer of $Q_w(\pi)$. \square

C. Experimental Setup

This section contains details of parameter tuning and data preprocessing for each set of experiments in Section 5 of the paper.

C.1. Partial AUC in $[0, \beta]$ for Ranking Applications

Cheminformatics. For SVM_{pAUC}/SVM_{AUC}, the error tolerance ϵ was set to 10^{-4} and C was selected from $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$ via 5-fold cross-validation on the training sample (based on highest average value of pAUC/AUC). For ASVM, the parameters μ and τ were chosen from $\{0.001, 0.005, 0.01, 0.05, 0.1\}$ and $\{0.0001, 0.0005, 0.001, 0.005, 0.01\}$, respectively, by similar cross-validation. For pAUCBoost, the iteration count was chosen similarly from the range $\{1, 2, \dots, 1000\}$. For the greedy heuristic algorithm, the parameter H (number of equally spaced intervals) was fixed to 100.

Information retrieval. For both SVM_{pAUC} and SVM_{AUC}, the regularization parameter C was selected from the range $\{0.001, 0.01, 0.1, 1, 10\}$ using the validation set (based on highest value of pAUC/AUC). In these experiments, the error tolerance parameter ϵ was set to a higher value 0.1, due to long running times with smaller values.¹

¹TD2004 (part of the LETOR 2.0 collection) was obtained from <http://research.microsoft.com/en-us/um/beijing/projects/letor/Letor2.0/dataset.aspx>; TREC10 was obtained from http://www.yisongyue.com/datasets/TREC_INDRI.tar.gz.

C.2. Partial AUC in $[0, \beta]$ for Protein-protein Interaction Prediction.

Algorithm parameters were chosen similarly to the cheminformatics data set (using a validation set). Continuous features were normalized to zero mean and unit variance; discrete features were scaled to $[0, 1]$. In this application, it is natural to consider a transductive setting since the set of all protein pairs is known; thus when normalizing features and filling in missing feature values, we used statistics computed from the entire data set.²

C.3. Partial AUC in $[\alpha, \beta]$ for Medical diagnosis.

Algorithm parameters were chosen as for the cheminformatics data set. All features were normalized to zero mean and unit variance (using only the training set in each case).³

C.4. Run Time Analysis

These experiments were run on an Intel Xeon (2.13 GHz) machine with 12 GB RAM; the algorithm parameters C and ϵ for these experiments were set to 10 and 0.1, respectively.

²The protein-protein interaction prediction data set was obtained from http://www.cs.cmu.edu/~qyj/papers_sulp/proteins05_pages/feature-download.html.

³The KDD Cup 08 breast cancer detection data set was obtained from <http://www.sigkdd.org/kddcup/>; since the test data provided here does not contain class labels, we used only the training data provided, splitting it randomly into multiple train-test splits.