Finding the Most Violated Constraint for CROC

Ayush Agarwal and Utkarsh Agarwal

1 Introduction

The area under **ROC** curve is a widely used performance measure but many applications don't concern themselves with the complete area under **ROC** . For example, in drug discovery problem only the top results matter and the application in not concerned as to how the model ranks the bottom results. So we need to design a method which focuses only on the top results. Luckily there are measures like **pAUC** and **CROC** (Concentrated **ROC**) which focus on optimising the top results. In this document, we try to model CROC as performance measure in struct SVM formulation.

2 Preliminaries

Let **X** be an instance space, and D_+, D_- be the probability distributions for positive and negative instances. Consider a training sample $S = (S_+,\ S_-)$ consisting of m positive instances $S_+ = (x_1^+, x_2^+, ...x_m^+)$ drawn iid according to D_+ and n negative instances $S_- = (x_1^-, x_2^-, ...x_n^-)$ drawn iid according to D_-. The end goal is to learn a scoring function $f : X \to \mathbb{R}$ which has good performance in terms of CROC.

2.1 CROC

Now we will define TPR and FPR in continuous as well as discrete case.
For continuous distribution:

True Positive Rate(**TPR**) of a classfier $sign(f(x) - t)$ is the probability that it correctly classifies a random positive instance from D_+.

$$TPR(f, t) = \mathbf{P}_{x^+ \sim D_+}[f(x^+) > t]$$

Similarly False Positive Rate (**FPR**) is the probability that the classifier completely misclassify a random negative instance as positive.

$$FPR(f, t) = \mathbf{P}_{x^- \sim D_-}[f(x^-) > t]$$

CROC is then defined as the plot of TPR_f(t) and $g(FPR_f(t))$ for different values of t where $g : [0, 1] \to [0, 1]$ is the CROC concentration function. Therefore,

$$CROC_f(g) = \int_0^1 TPR_f(FPR_f^{-1}(u))d(g(u))$$

Note: g must be such that $\int_0^1 d(g(x)) = 1$.

Empirical CROC:

$$TPR_f(t) = \frac{1}{m}\sum_{i=1}^{m}[f(x^+) > t]$$
$$FPR_f(t) = \frac{1}{n}\sum_{j=1}^{n}[f(x^-) > t]$$
$$CROC_f(g) = \frac{1}{c_1^n}\sum_{j=1}^{n} TPR_f(f(x^-_{(j)_f}))\left(g\left(\frac{j}{n}\right) - g\left(\frac{j-1}{n}\right)\right)$$
$$CROC_f(g) = \frac{1}{mc_1^n}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{1}(f(x_i^+) > f(x^-_{(j)_f}))\left(g\left(\frac{j}{n}\right) - g\left(\frac{j-1}{n}\right)\right)$$

Partial CROC:

$$pCROC_f(g) = \int_{\alpha}^{\beta} TPR_f(FPR_f^{-1}(u))d(g(u))$$

Empirical partial CROC(normalised):

$$pCROC_f(g) = \frac{1}{mc_\alpha^\beta}\sum_{i=1}^{m}\left[(j_\alpha - n\alpha)\mathbf{1}(f(x_i^+) > f(x^-_{(j^\alpha)_f}))\left(g\left(\frac{j_\alpha}{n}\right) - g\left(\frac{j_\alpha - 1}{n}\right)\right)\right.$$
$$\left. + \sum_{j=j_\alpha+1}^{\beta}\mathbf{1}(f(x_i^+) > f(x^-_{(j)_f}))\left(g\left(\frac{j}{n}\right) - g\left(\frac{j-1}{n}\right)\right) + (n\beta - j_\beta)\mathbf{1}(f(x_i^+) > f(x^-_{(j_\beta+1)_f}))\left(g\left(\frac{j_\beta+1}{n}\right) - g\left(\frac{j_\beta}{n}\right)\right)\right] \tag{1}$$

where $x^-_{(j)_f}$ is the j-th ranked negative point according to f and c_a^b is $\sum_{j=a}^{b}\left(g\left(\frac{j}{n}\right) - g\left(\frac{j-1}{n}\right)\right)$ the normalsation constant.

3 Structural SVM approach for optimising area under CROC

This section is based on [1].
We now cast our problem of optimising the area under CROC into structural SVM framework.
We will be using matrix $\pi = [\pi_{ij}]$ to denote the relative ordering among m positive and n negative examples belonging to S. Building on the approach of (Joachims, 2005) [3], we define a joint feature map

$$\phi : (X^m \times X^n) \times \Pi_{m,n} \to \mathbb{R}^d \quad \text{as}$$

$$\phi(S, \pi) = \frac{1}{mc_1^n}\sum_{i=1}^{m}\sum_{j=1}^{n}(1 - \pi_{ij})(x_i^+ - x_j^-)$$

We define CROC loss as

$$\triangle_{CROC}(\pi^*, \pi) = \frac{1}{mc_1^n}\sum_{i=1}^{m}\sum_{j=1}^{n}\pi_{ij} * \gamma_j$$

where

$$\gamma_j = g\left(\frac{j}{n}\right) - g\left(\frac{j-1}{n}\right)$$

and γ_j is the weight associated with the j-th rank among the negatives. Note that g here must be a non-decreasing function for γ_j to remain non-negative, which is a requirement for the formulation of maximisation objective in Section 4.
The choice of ϕ above ensures that for any fixed w$\in \mathbb{R}^d$, maximizing $w^\top \phi$ over $\pi \in \Pi_{m,n}$ yields an ordering matrix consistent with the scoring fucntion $f(x) = w^\top x$. The problem of optimising now reduces to finding a $w \in \mathbb{R}^d$ for which the maximiser over $\pi \in \Pi_{m,n}$ of $w^\top \phi(S, \pi)$ has the highest CROC. This can be framed as the following optimisation problem:

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

s.t. $\forall \pi \in \Pi_{m,n}$:

$$w^\top(\phi(S, \pi^*) - \phi(S, \pi)) \geq \triangle_{CROC}(\pi^*, \pi) - \xi$$

which in turn can be written as

$$\frac{1}{mc_1^n} \sum_{i=1}^{m} \sum_{j=1}^{n} \pi_{ij} w^\top(x_i^+ - x_j^-) \geq \triangle_{CROC}(\pi^*, \pi) - \xi \qquad \text{(OP1)}$$

4 Efficient Algorithm for finding the Most Violated Constraint

This section is based on [1].
The crux of the problem lies in having a polynomial time algorithm for finding the most violated constraint in each iteration for the combinatorial optimization problem (over $\pi \in \Pi_{m,n}$), which is stated as follows:

$$\bar{\pi} = \operatorname*{argmax}_{\pi \in \Pi_{m,n}} Q_w(\pi) \qquad \text{(OP2)}$$

where

$$Q_w(\pi) = \triangle_{CROC}(\pi^*, \pi) - \frac{1}{mc_1^n} \sum_{i=1}^{m} \sum_{j=1}^{n} \pi_{ij} w^\top(x_i^+ - x_j^-)$$

Definition: $\Pi_{m,n}^w = \{ \pi \in \Pi_{m,n}^w \mid \forall i, j1 < j2 : \pi_{i,(j1)_w} \geq \pi_{i,(j2)_w} \}$
This is the set of all ordering matrices π in which the negatives are sorted in descending order of the scores $w^\top x_j^-$.
Assumption: The solution $\bar{\pi}$ to OP2 lies in $\Pi_{m,n}^w$.

So, the optimisation problem can be rephrased as:

$$\bar{\pi} = \operatorname*{argmax}_{\pi \in \Pi_{m,n}^w} Q_w(\pi) \qquad \text{(OP3)}$$

Lemma: $\qquad Q_w(\pi) = \widetilde{Q}_w(\pi) \qquad \forall \pi \in \Pi_{m,n}^w$
where

$$\widetilde{Q}_w(\pi) = \frac{1}{mc_1^n} \sum_{i=1}^{m} \sum_{j=1}^{n} \pi_{i,(j)_w} * \gamma_j - \frac{1}{mc_1^n} \sum_{i=1}^{m} \sum_{j=1}^{n} \pi_{i,(j)_w} w^\top(x_{i,(j)_w}^{\pm})$$

$\widetilde{Q}_w(\pi)$ is the same function as that of $Q_w(\pi)$ with the only difference being that the negative points are rearranged and are now sorted according to w. Note that for each $\pi \in \Pi_{m,n}$ there is a corresponding

map in $\Pi^w_{m,n}$ and hence the equality between $\widetilde{Q}_w(\pi)$ and $Q_w(\pi)$.
Now

$$\begin{aligned} Q_w(\pi) &= \triangle_{CROC}(\pi^*, \pi) - \frac{1}{mc_1^n}\sum_{i=1}^{m}\sum_{j=1}^{n} \pi_{ij} w^\top (x_i^+ - x_j^-) \\ \widetilde{Q}_w(\pi) &= \frac{1}{mc_1^n}\sum_{i=1}^{m}\sum_{j=1}^{n} \pi_{i,(j)_w} * \gamma_j - \frac{1}{mc_1^n}\sum_{i=1}^{m}\sum_{j=1}^{n} \pi_{i,(j)_w} w^\top (x^{\pm}_{i,(j)_w}) \\ \widetilde{Q}_w(\pi) &= \frac{1}{mc_1^n}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} \pi_{i,(j)_w}\left(\gamma_j - w^\top x^{\pm}_{i,(j)_w}\right)\right] \end{aligned}$$

where $x^{\pm}_{ij} = (x_i^+ - x_j^-)$ and $(j)_w$ refers to the j-th ranked negative in S_- according to w.
Finally, we have

$$\bar{\pi} = \operatorname*{argmax}_{\pi \in \Pi^w_{m,n}} \widetilde{Q}_w(\pi) \qquad \text{(OP4)}$$

This problem is much easier as now we have sorted the negatives and in any range of ranks we will always get the same set of negatives. We solve a relaxation of OP4 where $\bar{\pi}$ is optimised over all $\pi \in \{0,1\}^{m\times n}$

$$\bar{\pi} = \operatorname*{argmax}_{\pi \in \{0,1\}^{m\times n}} \widetilde{Q}_w(\pi) \qquad \text{(OP5)}$$

OP5 yields

$$\bar{\pi}_{i,(j)_w} = \mathbf{1}(w^\top x^{\pm}_{i,(j)_w} \leq \gamma_j) \qquad (1)$$

Theorem1: The ordering matrix defined by Eq. (1) lies in $\Pi^w_{m,n}$.

Proof. Let us assume on the contrary that the ordering is not in $\Pi^w_{m,n}$.
Thus, $\exists\, j_1, j_2$ s.t. $j_1 < j_2$ and $\bar{\pi}_{i,(j_1)_w} < \bar{\pi}_{i,(j_2)_w}$ for some $i \in [m]$.

$$\begin{aligned} &\Rightarrow \bar{\pi}_{i,(j_1)_w} = 0 \text{ and } \bar{\pi}_{i,(j_2)_w} = 1 \\ &\Rightarrow w^\top x^{\pm}_{i,(j_1)_w} > \gamma_{j_1} \text{ and } w^\top x^{\pm}_{i,(j_2)_w} \leq \gamma_{j_2} \\ &\Rightarrow w^\top (x_i^+ - x^-_{(j_1)_w}) - \gamma_{j_1} > w^\top (x_i^+ - x^-_{(j_2)_w}) - \gamma_{j_2} \\ &\Rightarrow \gamma_{j_2} - \gamma_{j_1} > w^\top (x^-_{(j_1)_w} - x^-_{(j_2)_w}) \end{aligned}$$

Since $j_1 < j_2$, $\gamma_{j_1} > \gamma_{j_2}$(Proof in Extras(5)) but $w^\top x^-_{(j_1)_w} \geq w^\top x^-_{(j_2)_w}$.

$$\Rightarrow\Leftarrow \text{ Contradiction}$$

Hence the ordering lies in $\Pi^w_{m,n}$.

We can extend the algorithm when the CROC function is piecewise defined by optimising area along the lines of partial AUC for each definition of the function.

5 Extras

Let $g(x)$ be a continuous non linear concave function used for CROC. The weight γ_j associated with x_j^- is defined as

$$\gamma_j = g\left(\frac{j}{n}\right) - g\left(\frac{j-1}{n}\right)$$

Claim : γ is a decreasing function i.e. $\gamma_j < \gamma_{j-1}$.
Proof:

$$\begin{aligned}
\Rightarrow \quad & \gamma_{j-1} = g\left(\frac{j-1}{n}\right) - g\left(\frac{j-2}{n}\right) \\
\Rightarrow \quad & \gamma_j = g\left(\frac{j}{n}\right) - g\left(\frac{j-1}{n}\right) \\
\Rightarrow \quad & \gamma_{j-1} - \gamma_j = 2g\left(\frac{j-1}{n}\right) - \left(g\left(\frac{j-2}{n}\right) + g\left(\frac{j}{n}\right)\right)
\end{aligned}$$

Since g is a concave function, by Jensen's inequality we have

$$\begin{aligned}
& g\left(\frac{j-1}{n}\right) > \frac{1}{2}\left[g\left(\frac{j-2}{n}\right) + g\left(\frac{j}{n}\right)\right] \\
\Rightarrow \quad & 2g\left(\frac{j-1}{n}\right) > g\left(\frac{j-2}{n}\right) + g\left(\frac{j}{n}\right) \\
\Rightarrow \quad & \gamma_{j-1} > \gamma_j
\end{aligned}$$

Hence proved.

6 Structural SVM approach for optimising area under partial CROC

As above we define feature map as

$$\phi : (X^m \times X^n) \times \Pi_{m,n} \to \mathbb{R}^d \quad \text{as}$$

$$\phi(S, \pi) = \frac{1}{mc_\alpha^\beta} \sum_{i=1}^{m} \sum_{j=1}^{n} (1 - \pi_{ij})(x_i^+ - x_j^-)$$

Partial CROC loss is

$$\triangle_{pCROC}\,(\pi^*, \pi) = \frac{1}{mc_\alpha^\beta} \sum_{i=1}^{m} \left[(j_\alpha - n\alpha)\pi_{i,(j_\alpha)_\pi}\gamma_{j_\alpha} + \sum_{j=j_\alpha+1}^{\beta} \pi_{i,(j)_\pi}\gamma_j + (n\beta - j_\beta)\pi_{i,(j_\beta+1)_\pi}\gamma_{j_\beta+1}\right] \quad (2)$$

where

$$\gamma_j = g\left(\frac{j}{n}\right) - g\left(\frac{j-1}{n}\right)$$

and γ_j is the weight associated with the j-th rank among the negatives and $(j)_\pi$ denotes the index of the negative in S_- ranked in the jth position. Note that g here must be a non-decreasing function for γ_j to remain non-negative.
The problem of optimising now reduces to finding a $w \in \mathbb{R}^d$ for which the maximiser over $\pi \in \Pi_{m,n}$ of $w^\top \phi(S, \pi)$ has the highest partial CROC. This can be framed as the following optimisation problem:

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

s.t. $\forall \pi \in \Pi_{m,n}$:

$$w^\top(\phi(S, \pi^*) - \phi(S, \pi)) \geq \triangle_{pCROC}(\pi^*, \pi) - \xi$$

which in turn can be written as

$$\frac{1}{mc_\alpha^\beta}\sum_{i=1}^{m}\sum_{j=1}^{n}\ \pi_{ij}w^\top(x_i^+ - x_j^-) \geq \triangle_{pCROC}(\pi^*,\pi) - \xi \qquad \text{(OP6)}$$

7 Efficient Algorithm for finding the Most Violated Constraint

This section is based on [1].
We will try to design a polynomial time algorithm for finding the most violated constraint in each iteration for the combinatorial optimization problem (over $\pi \in \Pi_{m,n}$), which is stated as follows:

$$\bar{\pi} = \operatorname*{argmax}_{\pi\in\Pi_{m,n}} Q'_w(\pi) \qquad \text{(OP7)}$$

where

$$Q'_w(\pi) \ = \ \triangle_{pCROC}(\pi^*,\pi) \ - \ \frac{1}{mc_\alpha^\beta}\sum_{i=1}^{m}\sum_{j=1}^{n}\ \pi_{ij}w^\top(x_i^+ - x_j^-)$$

Assumption: The solution $\bar{\pi}$ to OP7 lies in $\Pi^w_{m,n}$.

So, the optimisation problem can be rephrased as:

$$\bar{\pi} = \operatorname*{argmax}_{\pi\in\Pi^w_{m,n}} Q'_w(\pi) \qquad \text{(OP8)}$$

Lemma: $$Q'_w(\pi) \ = \widetilde{Q'}_w\ (\pi) \qquad\qquad \forall\pi\in\Pi^w_{m,n}$$
where

$$\widetilde{Q'}_w\ (\pi) \ = \ \frac{1}{mc_\alpha^\beta}\sum_{i=1}^{m}\Bigg[(j_\alpha - n\alpha)\pi_{i,(j_\alpha)_w}\gamma_{j_\alpha} + \sum_{j=j_\alpha+1}^{\beta}\pi_{i,(j)_w}\gamma_j + (n\beta - j_\beta)\pi_{i,(j_\beta+1)_w}\gamma_{j_\beta+1}\Bigg] - \frac{1}{mc_\alpha^\beta}\sum_{i=1}^{m}\sum_{j=1}^{n}\ \pi_{i,(j)_w}w^\top(x^\pm_{i,(j)_w}) \quad (3)$$

The proof is identical to the proof of a similar lemma stated earlier.
Now

$$\begin{aligned}
Q'_w(\pi) \ &= \ \triangle_{pCROC}(\pi^*,\pi) \ - \ \frac{1}{mc_\alpha^\beta}\sum_{i=1}^{m}\sum_{j=1}^{n}\ \pi_{ij}w^\top(x_i^+ - x_j^-) \\
\widetilde{Q'}_w\ (\pi) \ &= \ \frac{1}{mc_\alpha^\beta}\sum_{i=1}^{m}\Bigg[(j_\alpha - n\alpha)\pi_{i,(j_\alpha)_w}\gamma_{j_\alpha} + \sum_{j=j_\alpha+1}^{\beta}\pi_{i,(j)_w}\gamma_j + (n\beta - j_\beta)\pi_{i,(j_\beta+1)_w}\gamma_{j_\beta+1}\Bigg] \\
&- \frac{1}{mc_\alpha^\beta}\sum_{i=1}^{m}\sum_{j=1}^{n}\ \pi_{i,(j)_w}w^\top(x^\pm_{i,(j)_w}) \\
&= \ \frac{1}{mc_\alpha^\beta}\sum_{i=1}^{m}\Bigg[-\sum_{j=1}^{j_\alpha-1}\ \pi_{i,(j)_w}w^\top(x^\pm_{i,(j)_w}) + \pi_{i,(j_\alpha)_w}\left[(j_\alpha - n\alpha)\gamma_{(j_\alpha)} - w^\top(x^\pm_{i,(j_\alpha)_w})\right] \\
&+ \sum_{j=j_\alpha+1}^{\beta}\pi_{i,(j)_w}(\gamma_j - w^\top(x^\pm_{i,(j)_w})) + \pi_{i,(j_\beta+1)_w}\left[(n\beta - j_\beta)\gamma_{(j_\beta+1)} - \pi_{i,(j)_w}w^\top(x_{i,(j_\beta+1)_w}\right] \\
&- \sum_{j=j_\beta+2}^{n}\ \pi_{i,(j)_w}w^\top(x^\pm_{i,(j)_w})\Bigg]
\end{aligned}$$

where $x_{ij}^{\pm} = (x_i^+ - x_j^-)$ and $(j)_w$ refers to the j-th ranked negative in S_- according to w.
Finally, we have

$$\bar{\pi} = \underset{\pi \in \Pi_{m,n}^{w}}{\operatorname{argmax}} \ \widetilde{Q'}_w(\pi) \qquad \text{(OP9)}$$

This problem is much easier as now we have sorted the negatives and in any range of ranks we will always get the same set of negatives.
We have to look at the restricted space $\Pi_{m,n}^{w}$ as the solution will not lie in $\{0,1\}^{(m,n)}$. This can be seen as the coefficients of $\pi_{i,(j)_w}$ are not same for all $j \in \{1,n\}$. As we will see we can still find the values of π matrix row wise.
π_i is of the form $\{1,1,1,.....,1,0,0,0,.....,0\}$ and k is the least index such that $\pi_{i,(k)_w} = 0$.

Algorithm 1 Find the Most Violated Constraint for Partial CROC

procedure MOST-VIOLATED(S_+,S_-, α, β, w)

  for $i = 1$ to m **do**

    Optimize over $k \in \{0,.....,j_\alpha - 1\}$

$$\pi^1_{i,(j)_w} = \begin{cases} \mathbf{1}(w^\top x^{\pm}_{i,(j)_w} \leq 0) & j \in \{1,...,j_\alpha - 1\} \\ \mathbf{0} & j \in \{j_\alpha,...,n\} \end{cases}$$

    Optimize over $k = j_\alpha$

$$\pi^2_{i,(j)_w} = \begin{cases} \mathbf{1} & j \in \{1,...,j_\alpha\} \\ \mathbf{0} & j \in \{j_\alpha + 1,...,n\} \end{cases}$$

    Optimize over $k \in \{j_\alpha + 1,.....,n\}$

$$\pi^3_{i,(j)_w} = \begin{cases} \mathbf{1} & j \in \{1,...,j_\alpha + 1\} \\ \mathbf{1}(w^\top x^{\pm}_{i,(j)_w} \leq \gamma_j) & j \in \{j_\alpha + 1,...,j_\beta\} \\ \mathbf{1}(w^\top x^{\pm}_{i,(j)_w} \leq (n\beta - j_\beta)\gamma_j) & j \in \{j_\beta + 1\} \\ \mathbf{1}(w^\top x^{\pm}_{i,(j)_w} \leq 0) & j \in \{j_\beta + 2,...,n\} \end{cases}$$

    $\pi_i^* = \underset{l \in 1,2,3}{argmax}(Q'_w(\pi^l_{i,(j)_w}))$

  Output : $\bar{\pi}$

8 $\text{SVM}^{\text{struct}}_{\text{pCROC}}$: Struct SVM formulation for partial CROC

This section is based on [2].
We will attempt to improve the bound on partial CROC than that in section 6.

8.1 Error upper bound formulation for $\text{SVM}^{\text{struct}}_{\text{pCROC}}$

Assumption: $n\alpha$ and $n\beta$ are integers.

Let us define $l^c_{ij}(w) = (c - w^\top(x^+_i - x^-_j))_+$ to be the c margin hinge loss function of $w \in \mathbb{R}^d$ and instance pair $(x^+_i, x^-_j)$for any $c \geq 0$.

Theorem2: Let $0 < \alpha < \beta \leq 1$. Then for any $w \in \mathbb{R}^d$ the smallest ξ satisfying the constraints of OP6 can be characterised as follows:

$$\xi = \xi_{pCROC} + \xi_{extra}$$

where

$$\xi_{pCROC} \leq \frac{1}{mc^\beta_\alpha} \sum_{i=1}^{m} \Big[\sum_{j=1}^{j_\alpha} l^{(0)}_{i,(j)_w}(w) + \sum_{j=j_\alpha+1}^{j_\beta} l^{(\gamma_j)}_{i,(j)_w}(w) \Big]$$
$$\xi_{extra} = \frac{1}{mc^\beta_\alpha} \sum_{i=1}^{m} \sum_{j=j_\beta+1}^{n} l^{(0)}_{i,(j)_w}(w)$$

Proof:
The smallest $\xi \geq 0$ satisyfing OP6 can be written as:

$$\xi = \frac{1}{mc^\beta_\alpha} \sum_{i=1}^{m} \Bigg[- \sum_{j=1}^{j_\alpha} \pi_{i,(j)_w} w^\top (x^\pm_{i,(j)_w}) + \sum_{j=j_\alpha+1}^{\beta} \pi_{i,(j)_w} (\gamma_j - w^\top (x^\pm_{i,(j)_w})) - \sum_{j=j_\beta+1}^{n} \pi_{i,(j)_w} w^\top (x^\pm_{i,(j)_w}) \Bigg]$$

We bound each $\pi_{i,(j)_w}$ independently with the hinge loss function. Thus we get an upper bound on ξ_{pCROC} and an extra non-negative term ξ_{extra}.

9 $\text{SVM}^{\text{tight}}_{\text{pCROC}}$ formulation

This section is based on [2].
Loss for partial CROC when $n\alpha$ and $n\beta$ are integers is

$$\triangle_{pCROC}(\pi^*, \pi) = \frac{1}{mc^\beta_\alpha} \sum_{i=1}^{m} \sum_{j=j_\alpha+1}^{\beta} \pi_{i,(j)_\pi} \gamma_j$$

Let N_β be the set of all subsets of S_- such that $|S_-| = j_\beta$. We now redefine loss as

$$\triangle_{pCROC}(\pi^*, \pi, N_\beta) = \frac{1}{mc^\beta_\alpha} \max_{z \in N} \sum_{i=1}^{m} \sum_{x^-_j \in z} \mathbf{1}(f(x^+_i) < f(x^-_j))\gamma_j \tag{4}$$

Theorem3: $\triangle_{pCROC}(\pi^*, \pi, N_\beta)$ is equivalent to $\triangle_{pCROC}(\pi^*, \pi)$
Proof: Consider only one fixed positive x^+_i. The partial CROC loss function is only dependent on its relative rank wrt the negatives ranked from j_α to j_β. Since γ_j is a decreasing continuous function of j, if we consider all possible reorderings of negatives $\mathbf{1}(f(x^+_i) < f(x^-_j))\gamma_j$ will equate the analogous parameter in loss function when negatives are sorted in decreasing order by f. Hence the redefinition is equivalent to the original definition.

Theorem4: The maximum in the equation 5 is attained for top j_β negatives ranked by f(in descending order).
Proof: Consider a fixed i.
Now consider negatives as ordered by f in decreasing order with score of each negative as $\mathbf{1}(f(x_i^+) < f(x_j^-))\gamma_j$. The score list will be of the form $\gamma_1, \gamma_2, ..., \gamma_k, 0, 0, ..., 0$. Now we have to select $j_\beta - j_\alpha$ elements from this which will form the summation term. Since γ_j decreases with increase in j, the score list is sorted in decreasing order, thus we will achieve a maximum when we pick the first j_β elements or, in other words when we select the top j_β ranked negatives.

Based on the loss function for pCROC in equation 5 we devise a new SVM formulation which gives a tighter upper bound on pCROC loss. For a given subset of negative instances $z = \{x_{k1}^-, x_{k2}^-, ..., x_{kj_\beta}\} \in N_\beta$ we define the joint feature map

$$\phi_z : (X^m \times X^n) \times \Pi_{m,j_\beta} \to \mathbb{R}^d \quad \text{as}$$

$$\phi_z(S, \pi) = \frac{1}{mc_\alpha^\beta} \sum_{i=1}^{m} \sum_{j=1}^{j_\beta} (1 - \pi_{ij})(x_i^+ - x_j^-)$$

The loss function is

$$\triangle_{pCROC}(\pi^*, \pi) = \frac{1}{mc_\alpha^\beta} \sum_{i=1}^{m} \sum_{j=j_\alpha+1}^{\beta} \pi_{i,(j)_\pi} \gamma_j$$

Thus $\text{SVM}_{\text{pCROC}}^{\text{tight}}$ solves the following convex optimisation problem:

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

s.t. $\forall z \in N_\beta, \forall \pi \in \Pi_{m,j_\beta}$:

$$w^\top(\phi_z(S, \pi^*) - \phi_z(S, \pi)) \geq \triangle_{pCROC}(\pi^*, \pi) - \xi$$

where $C > 0$ is the regularization parameter.
Theorem5: The maximum value of ξ is attained for (z^*, π^*) where z^* is a subset of N_β containing top j_β negatives ranked in descending order by $w^\top x$.
Proof: For a fixed ordering matrix π the only term in the expansion of ξ that depends on the set $z = \{x_{k1}^-, x_{k2}^-, ..., x_{kj_\beta}\}$ is

$$\frac{1}{mc_\alpha^\beta} \sum_{i=1}^{m} \sum_{j=1}^{j_\beta} [\pi_{i,(j)_w}] w^\top (x_{(kj)}^-)$$

Since $\pi_{i,(j)_w}$ is fixed it depends only on the summation of scores of z by w, it is trivially maximum for z^*.
Theorem6: The smallest ξ satisfying the constraints of for $0 < \alpha < \beta \leq 1$ is ξ_{pCROC}.
Proof: The smallest ξ satisfying the constraints can be expanded as

$$\begin{aligned}
\xi &\geq \triangle_{pCROC}(\pi^*, \pi) - w^\top(\phi_z(S, \pi) - \phi_z(S, \pi^*)) \\
&\geq \frac{1}{mc_\alpha^\beta} \sum_{i=1}^{m} \sum_{j=j_\alpha+1}^{\beta} \pi_{i,(j)_\pi} \gamma_j - \frac{1}{mc_\alpha^\beta} w^\top \left[\sum_{i=1}^{m} \sum_{j=1}^{j_\beta} \pi_{ij}(x_i^+ - x_j^-) \right] \\
&\geq \frac{1}{mc_\alpha^\beta} \sum_{i=1}^{m} \left[-\sum_{j=1}^{j_\alpha} \pi_{i,(j)_w} w^\top (x_{i,(j)_w}^\pm) + \sum_{j=j_\alpha+1}^{\beta} \pi_{i,(j)_w}(\gamma_j - w^\top(x_{i,(j)_w}^\pm)) \right]
\end{aligned}$$

Thus ξ is essentially ξ_{pCROC}.

10 Most violated constraint for $\mathbf{SVM}^{\mathbf{tight}}_{\mathbf{pCROC}}$

By Theorem 5, the maximum value of ξ is attained for z^* which contains the top j_β ranked negatives. $z*$ can be easily found by sorting the negatives in deceasing order of their scores by w and picking up the top j_β elements. For a fixed z^*, the problem reduces to finding the ordering matrix $\pi*$ on a reduced domain Π_{m,j_β}, which is equivalent to finding the most violated constraint in $\mathrm{SVM}^{\mathrm{Struct}}_{\mathrm{pCROC}}$ with modified input $(S^+, z^*, \alpha, \beta, w)$.

References

[1] H. Narasimhan and S. Agarwal. A structural SVM based approach for optimizing partial AUC. In ICML, 2013,

[2] H. Narasimhan and S. Agarwal. SVMpAUC-tight A New Support Vector Method for Optimizing Partial AUC Based on a Tight Convex Upper Bound In KDD, 2013,

[3] Joachims, T. A support vector method for multivariate performance measures. In Proceedings of the 22nd International Conference on Machine Learning, 2005.