We provide missing proofs of theorems and extensions that were excluded from the main body of the paper due to space constraints.

12 Regret for SumLoss

12.1 Proof of Lemma 2

Proof. For any $p \in \Delta$, we have $\ell_i \cdot p = \sum_{j=1}^{2^m} p_j \ (\sigma_i \cdot r_j) = \sigma_i \cdot (\sum_{j=1}^{2^m} p_j r_j) = \sigma_i \cdot E_r[r]$, where the expectation is taken w.r.t. p (p_j is the jth component of p). By dot product rule between 2 vectors, $l_i \cdot p$ is minimized when ranking of objects according to σ_i and expected relevance of objects are in opposite order. That is, the object with highest expected relevance is ranked 1 and so on. Formally, $l_i \cdot p$ is minimized when $E_r[r(\sigma_i^{-1}(1)] \geq E_r[r(\sigma_i^{-1}(2)] \geq \ldots \geq E_r[r(\sigma_i^{-1}(m)]$.

Thus, for action i, probability cell is defined as $C_i = \{p \in \Delta : \sum_{j=1}^{2^m} p_j = 1,\ E_r[r(\sigma_i^{-1}(1)] \geq E_r[r(\sigma_i^{-1}(2)] \geq \ldots \geq E_r[r(\sigma_i^{-1}(m)]\}$. Note that, $p \in C_i$ iff action i is optimal w.r.t. p. Since C_i is obviously non-empty and it has only 1 equality constraint (hence $2^m - 1$ dimensional), action i is Pareto optimal.

The above holds true for all learner's actions i. ☐

12.2 Proof of Lemma 3

Proof. From Lemma 2, we know that every one of learner's actions is Pareto-optimal and C_i, associated with action σ_i, has structure $C_i = \{p \in \Delta : \sum_{j=1}^{2^m} p_j = 1,\ E_r[r(\sigma_i^{-1}(1)] \geq E_r[r(\sigma_i^{-1}(2)] \geq \ldots > E_r[r(\sigma_i^{-1}(m)]\}$.

Let $\sigma_i^{-1}(k) = a,\ \sigma_i^{-1}(k+1) = b$. Let it also be true that $\sigma_j^{-1}(k) = b,\ \sigma_j^{-1}(k+1) = a$ and $\sigma_i^{-1}(n) = \sigma_j^{-1}(n),\ \forall n \neq \{k,\ k+1\}$. Thus, objects in $\{\sigma_i, \sigma_j\}$ are same in all places except in a pair of consecutive places where the objects are interchanged.

Then, $C_i \cap C_j = \{p \in \Delta : \sum_{j=1}^{2^m} p_j = 1,\ E_r[r(\sigma_i^{-1}(1)] \geq \ldots \geq E_r[r(\sigma_i^{-1}(k)] = E_r[r(\sigma_i^{-1}(k+1)] \geq \ldots \geq E_r[r(\sigma_i^{-1}(m)]\}$. Hence, there are two equalities in the non-empty set $C_i \cap C_j$ and it is an $(2^m - 2)$ dimensional polytope. Hence condition of Definition 4 holds true and $\{\sigma_i, \sigma_j\}$ are neighboring actions pair.

☐

12.3 Proof of Theorem 4

Proof. We will explicitly show that local observability condition fails by considering the case when number of objects is $m = 3$. Specifically, action pair $\{\sigma_1,\ \sigma_2\}$, in Table 1 are neighboring actions, using Lemma 3 . Now every other action $\{\sigma_3, \sigma_4, \sigma_5, \sigma_6\}$ either places object 2 at top or object 3 at top. It is obvious that the set of probabilities for which $E_r[r(1)] \geq E_r[r(2)] = E_r[r(3)]$ cannot be a subset of any C_3, C_4, C_5, C_6. From Def. 4, the neighborhood action set of actions $\{\sigma_1, \sigma_2\}$ is precisely σ_1 and σ_2 and contains no other actions. By definition of signal matrices $S_{\sigma_1},\ S_{\sigma_2}$ and entries $\ell_1,\ \ell_2$ in Table 1 and 2, we have,

$$S_{\sigma_1} = S_{\sigma_2} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \qquad (8)$$
$$\ell_1 - \ell_2 = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 \end{bmatrix}.$$

It is clear that $\ell_1 - \ell_2 \notin Col(S_{\sigma_1}^{\top})$. Hence, Definition 5 fails to hold. ☐

13 Efficient Algorithm for Obtaining Regret

13.1 Proof of Lemma 8

Proof. We can write $\hat{r}_t = \sum_{j=1}^{m} r_{i_j}(j) e_j$, where e_j is the standard basis vector along coordinate j. Then $E_{i_1,\ldots,i_m}(\hat{r}_t) = \sum_{j=1}^{m} E_{i_j}(r_{i_j}(j) e_j) = \sum_{j=1}^{m} \sum_{k=1}^{t} \dfrac{r_k(j) e_j}{t} = r_{1:t}^{avg}$. ☐

13.2 Proof of Theorem 7

Proof. The proof for the top-1 feedback needs a careful look at the analysis of FTPL when we divide time into phases/blocks.

FTPL with blocking. Instead of top-1 feedback, assume that at each round, after learner reveals his action, the full relevance vector is revealed to the learner. Then an $O(\sqrt{T})$ expected regret for *SumLoss* can be obtained by applying FTPL (Follow the perturbed leader), in the following manner.

At end of every round t, the full relevance vector generated by the adversary is revealed. The relevance vectors are accumulated as $r_{1:t} = r_{1:t-1} + r_t$, where $r_{1:s} = \sum_{i=1}^{s} r_i$. A learner's action (permutation) for round $t+1$ is generated by solving $M(r_{1:t} + p_t)$, where $p_t \in [0, \frac{1}{\epsilon}]^m$ (uniform distribution) and ϵ is algorithmic (randomization) parameter. It should be noted that $M(y) = \underset{\sigma}{\text{argmin}}\ \sigma \cdot y$ is simply sorting of y since $f(\sigma) = \sigma$ is a monotone function as defined in Sec. 3 .

The key idea is that FTPL implicitly maintains a distribution over $m!$ actions (permutations) at beginning of each round, by randomly perturbing the scores of only m objects: score of each object is sum of (deterministic) accumulated relevance so far and (random) uniform value from $[0, \frac{1}{\epsilon}]$. Thus, it bypasses having to maintain explicit weight on each of $m!$ arms, which is computationally prohibitive. This key property which introduces efficiency in our algorithm is in contrast to the general algorithms based on exponential weights, which have to maintain explicit weights, based on accumulated history, on each action and randomly select an action based on weights.

Now let us look at a variant of the full information problem. The (known) time horizon T is divided into K blocks, i.e., $\{B_1, \ldots, B_K\}$, of equal size $\lfloor T/K \rfloor$. Here, $B_i = \{(i-1)(T/K)+1, (i-1)(T/K)+2, (i-1)T/K+3, \ldots, i(T/K)\}$. While operating in a block, the relevance vectors within the block are accumulated, but not used to generate learner's actions like in the full information version. Assume at the start of block B_i, there was some relevance vector r^i. Then at each time point in the block, a fresh $p \in [0, \eta]^m$ is sampled and $M(r^i + p)$ is solved to generate permutation for next time point. At the end of a block, the average of the accumulated relevance vectors (r^{avg}) for the block is used for updation, as $r^i + r^{avg}$, to get r^{i+1} for the next block. The process is repeated for each block. At the beginning of the first block, $r^1 = \{0\}^m$.

Formally, let the FTPL have an implicit distribution ρ_i (over the permutations) at the beginning of block B_i. That is $\rho_i \in \Delta$, where Δ is the probability simplex over $m!$ actions. Sampling a permutation using ρ_i at each time point of the block B_i means sampling a fresh $p \in [0, \eta]^m$ at every time point t and solving $M(s_{1:(i-1)} + p)$, where $s_{1:(i-1)} = \sum_{j=1}^{i-1} s_j$ and s_j is the average of relevance vectors of block B_j. Note that the distribution ρ_i is a fixed, deterministic function of the vectors $s_1, \ldots, s_{i-1}$.

Since action σ_t, for $t \in B_k$, is generated according to distribution ρ_k (we will denote this as $\sigma_t \sim \rho_k$), and in block k, distribution ρ_k is fixed, we have

$$E_{\sigma_t \sim \rho_k}[\sum_{t \in [B_k]} SumLoss(\sigma_t, r_t)] =$$
$$\sum_{t \in B_k} \rho_k \cdot [SumLoss(\sigma_1, r_t), \ldots, SumLoss(\sigma_{m!}, r_t)].$$

(dot product between 2 vectors of length $m!$).

Thus, the total expected loss of this variant of the full information problem is:

$$\sum_{t=1}^{T} E_{\sigma_t \sim \rho_k}[SumLoss(\sigma_t, r_t)] =$$
$$\sum_{k=1}^{K} E_{\sigma_t \sim \rho_k}[\sum_{t \in B_k} SumLoss(\sigma_t, r_t)] \tag{9}$$
$$= \sum_{k=1}^{K} \sum_{t \in B_k} \rho_k \cdot [SumLoss(\sigma_1, r_t), \ldots, SumLoss(\sigma_{m!}, r_t)]$$
$$= \sum_{k=1}^{K} \sum_{t \in B_k} \rho_k \cdot [\sigma_1 \cdot r_t, \ldots, \sigma_{m!} \cdot r_t)] \text{ By defn. of SumLoss}$$
$$= \frac{T}{K} \sum_{k=1}^{K} \rho_k \cdot [\sigma_1 \cdot s_k, \ldots, \sigma_{m!} \cdot s_k]$$
$$= \frac{T}{K} \sum_{k=1}^{K} E_{\sigma_k \sim \rho_k}[SumLoss(\sigma_k, s_k)]$$
$$= \frac{T}{K} E_{\sigma_1 \sim \rho_1, \ldots, \sigma_K \sim \rho_K} \sum_{k=1}^{K} SumLoss(\sigma_k, s_k) \tag{10}$$

where $s_k = \sum_{t \in B_k} \frac{r_t}{T/K}$. Note that, at end of every round $k \in [K]$, ρ_k is updated to ρ_{k+1} by feeding $s_{1:k}$ to FTPL. By the regret bound of FTPL, for K rounds of full information problem, with $\epsilon = \sqrt{D/RAK}$, we have:

$$\begin{aligned} & E_{\sigma_1 \sim \rho_1, \ldots, \sigma_K \sim \rho_K} \sum_{k=1}^{K} SumLoss(\sigma_k, s_k) \\ & \leq \min_{\sigma} \sum_{k=1}^{K} SumLoss(\sigma, s_k) + 2\sqrt{DRAK} \\ & = \min_{\sigma} \sum_{k=1}^{K} \sigma \cdot s_k + 2\sqrt{DRAK} \\ & = \min_{\sigma} \sum_{t=1}^{T} \sigma \cdot \frac{r_t}{T/K} + 2\sqrt{DRAK} \end{aligned} \tag{11}$$

where D, R, A are parameters dependent on the loss under consideration, that we will discuss and set later.

Now, since

$$\min_{\sigma} \sum_{t=1}^{T} \sigma \cdot \frac{r_t}{T/K} = \min_{\sigma} \frac{1}{T/K} \sum_{t=1}^{T} SumLoss(\sigma, r_t),$$

combining Eq. 9 and Eq. 11, we get:

$$\begin{aligned} & \sum_{t=1}^{T} E_{\sigma_t \in \rho_k}[SumLoss(\sigma_t, r_t)] \\ & \leq \min_{\sigma} \sum_{t=1}^{T} SumLoss(\sigma, r_t) + 2\frac{T}{K}\sqrt{DRAK}. \end{aligned} \tag{12}$$

FTPL with blocking and top-1 feedback. However, in our top-1 feedback model, the learner does not get to see the full relevance vector at each round. Thus, we form the unbiased estimator $\hat{s}_k$ of s_k, using Lemma 8. That is, at start of each block, we choose m time points uniformly at random, and at those time points, we output a random permutation which places each object, in turn, at top. At the end of the block, we form the relevance vector $\hat{s}_k$ which is the unbiased estimator of s_k. Note that using $\hat{s}_k$ instead of true s_k makes the distributions ρ_k themselves random. But significantly, ρ_k is dependent only on information received upto the beginning of block k and is independent of the information collected in the block. Thus, for block k, we have:

$$E_{\sigma_t \sim \rho_k(\hat{s}_1,\hat{s}_2,..,\hat{s}_{k-1})} \sum_{t \in [B_k]} SumLoss(\sigma_t, r_t)$$
$$= \frac{T}{K} E_{\sigma_k \sim \rho_k(\hat{s}_1,\hat{s}_2,..,\hat{s}_{k-1})} SumLoss(\sigma_k, s_k)$$
(From Eq. 9)
$$= \frac{T}{K} E_{\sigma_k \sim \rho_k(\hat{s}_1,\hat{s}_2,..,\hat{s}_{k-1})} E_{\hat{s}_k} SumLoss(\sigma_k, \hat{s}_k)$$
($\because$ SumLoss is linear in s and $\hat{s}_k$ is unbiased)
$$= \frac{T}{K} E_{\hat{s}_k} E_{\sigma_k \sim \rho_k(\hat{s}_1,\hat{s}_2,..,\hat{s}_{k-1})} SumLoss(\sigma_k, \hat{s}_k).$$

In the last step above, we crucially used the fact that, since random distribution ρ_k is independent of $\hat{s}_k$, the order of expectations is interchangeable. Taking expectation w.r.t. $\hat{s}_1, \hat{s}_2, .., \hat{s}_{k-1}$, we get,

$$E_{\hat{s}_1,\ldots,\hat{s}_{k-1}} E_{\sigma_t \sim \rho_k(\hat{s}_1,\hat{s}_2,..,\hat{s}_{k-1})} \sum_{t \in [B_k]} SumLoss(\sigma_t, r_t) =$$
$$\frac{T}{K} E_{\hat{s}_1,\ldots,\hat{s}_{k-1},\hat{s}_k} E_{\sigma_t \sim \rho_k(\hat{s}_1,\hat{s}_2,..,\hat{s}_{k-1})} SumLoss(\sigma_k, \hat{s}_k).$$

Thus,

$$\mathbb{E} \sum_{t=1}^{T} SumLoss(\sigma_t, r_t) = \mathbb{E} \sum_{k=1}^{K} \sum_{t \in [B_k]} SumLoss(\sigma_t, r_t)$$
$$\sum_{k=1}^{K} E_{\hat{s}_1,\ldots,\hat{s}_{k-1}} E_{\sigma_t \sim \rho_k(\hat{s}_1,\hat{s}_2,..,\hat{s}_{k-1})} \sum_{t \in [B_k]} SumLoss(\sigma_t, r_t)$$
$$\frac{T}{K} \sum_{k=1}^{K} E_{\hat{s}_1,\ldots,\hat{s}_{k-1},\hat{s}_k} E_{\sigma_t \sim \rho_k(\hat{s}_1,\hat{s}_2,..,\hat{s}_{k-1})} SumLoss(\sigma_k, \hat{s}_k)$$
$$= \frac{T}{K} E_{\hat{s}_1,\ldots,\hat{s}_K} \sum_{k=1}^{K} E_{\sigma_t \sim \rho_k(\hat{s}_1,\hat{s}_2,..,\hat{s}_{k-1})} SumLoss(\sigma_k, \hat{s}_k)$$

Now using Eq. 11, we can upper bound the last term above as

$$\leq \frac{T}{K} \{ E_{\hat{s}_1,\ldots,\hat{s}_K} [\min_{\sigma} \sum_{k=1}^{K} \sigma \cdot \hat{s}_k] + 2\sqrt{DRAK} \}$$
$$\leq \frac{T}{K} \{ \min_{\sigma} \sum_{k=1}^{K} \sigma \cdot s_k + 2\sqrt{DRAK} \}$$
(Jensen's Inequality)
$$\leq \min_{\sigma} \sum_{t=1}^{T} \sigma \cdot r_t + 2\frac{T}{K}\sqrt{DRAK}$$
$$= \min_{\sigma} \sum_{t=1}^{T} SumLoss(\sigma, r_t) + 2\frac{T}{K}\sqrt{DRAK}.$$

However, since in each block B_k, m rounds are reserved for exploration, where we do not draw σ_t from distribution ρ_k, the total expected loss is higher. Exploration leads to an extra regret of RmK, where R, as has been stated before, is an implicit parameter depending on the loss under consideration. The extra regret is because loss in each of the exploration rounds is at most R and there are a total of mK exploration rounds over all K blocks. Thus, overall regret is larger by RmK giving us:

$$E\left[\sum_{t=1}^{T} SumLoss(\sigma_t, r_t)\right] - \min_{\sigma} \sum_{t=1}^{T} SumLoss(\sigma, r_t)$$
$$\leq RmK + 2\frac{T}{K}\sqrt{DRAK}.$$

Now we optimize over K and set $K = (DA/R)^{1/3}(T/m)^{2/3}$, to get:

$$E\left[\sum_{t=1}^{T} SumLoss(\sigma_t, r_t)\right] \leq \min_{\sigma} \sum_{t=1}^{T} SumLoss(\sigma, r_t)$$
$$+ O(m^{1/3}R^{2/3}(DA)^{1/3}T^{2/3}) \tag{13}$$

Now, we recall the definitions of D, R and A from Kalai and Vempala [2005]: D is an upper bound on the ℓ_1 norm of vectors in learner's action space, R is an upper bound on the dot product of vectors in learner's and adversary's action space, and A is an upper bound on the ℓ_1 norm on vectors in adversary's action space. Thus, for $SumLoss$, we have

$$D = \sum_{i=1}^{m} \sigma(i) = O(m^2),$$
$$R = \sum_{i=1}^{m} \sigma(i) r(i) = O(m^2),$$
$$A = \sum_{i=1}^{m} r(i) = O(m).$$

Plugging in these values gives us Theorem 7. □

14 Regret Bounds for DCG and Prec@k

We deal with DCG first followed by Prec@k.

14.1 Extension of Results of SumLoss to DCG

We give pointers in the direction of proving the following results: a) Local observability condition fails to hold for DCG, b)The efficient algorithm of Sec.7 applies to DCG, with regret of $O(T^{2/3})$. Thus, the minimax regret of DCG is $\Theta(T^{2/3})$. All results are applicable to non-binary relevance vectors. The application of Algorithm 1 allows us to skip the proof of global observability, which is complicated for non-binary relevance vectors.

Let adversary be able to choose $r \in \{0, 1, \ldots, n\}^m$. Then, from definition of DCG in Sec.3 , it is clear DCG=$f(\sigma) \cdot g(r)$. $f(\sigma)$ and $g(r)$ has already been defined for DCG. Both are composed of m copies of univariate, monotonic, scalar valued function, where for $f(\cdot)$, it is monotonically decreasing whereas for $g(\cdot)$, it is increasing.

With slight abuse of notations, the loss matrix L implicitly means gain matrix, where entry in cell $\{i, j\}$ of L is $f(\sigma_i) \cdot g(r_j)$. The feedback matrix H remains the same. In Definition 1, learner action i is optimal if $\ell_i \cdot p \geq \ell_j \cdot p$, $\forall j \neq i$.

In Definition 2, the maximum number of distinct elements that can be in a row of H is $n + 1$. The signal matrix now becomes $S_i \in \{0, 1\}^{(n+1)\times 2^m}$, where $(S_i)_{k,\ell} = \mathbb{1}(H_{i,\ell} = k - 1)$.

14.1.1 Local Observability Fails

Since we are trying to establish a lower bound, it is sufficient to show it for binary relevance vectors.

In Lemma 2, proved for SumLoss, $\ell_i \cdot p$ equates to $f(\sigma) \cdot E_r[r]$. From definition of DCG, and from the structure and properties of $f(\cdot)$, it is clear that $\ell_i \cdot p$ is *maximized* under the same condition, i.e, $E_r[r(\sigma_i^{-1}(1)] \geq E_r[r(\sigma_i^{-1}(2)] \geq \ldots \geq E_r[r(\sigma_i^{-1}(m)]$. Thus, all actions are Pareto-optimal.

Careful observation of Lemma 3 shows that it is directly applicable to DCG, in light of extension of Lemma 2 to DCG.

Finally, just like in SumLoss, simple calculations with $m = 3$ and $n = 1$, in light of Lemma 2 and 3, show that local observability condition fails to hold.

We show the calculations:

$$S_{\sigma_1} = S_{\sigma_2} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\begin{aligned} \ell_{\sigma_1} =& [0, 1/2, 1/\log_2 3, 1/2 + 1/\log_2 3, 1, 3/2, \\ & 1 + 1/\log_2 3, 3/2 + 1/\log_2 3] \\ \ell_{\sigma_2} =& [0, 1/\log_2 3, 1/2, 1/2 + 1/\log_2 3, 1, 1 + 1/\log_2 3, \\ & 3/2, 3/2 + 1/\log_2 3] \end{aligned}$$

It is clear that $\ell_1 - \ell_2 \notin Col(S_{\sigma_1}^\top)$. Hence, Definition 5 fails to hold.

14.1.2 Implementation of the Efficient Algorithm

The only change in Algorithm 1 that allows extension to DCG with non-binary relevance is that relevance values will enter into the algorithm via the transformation $g^s(\cdot)$. That is, every component of relevance vector r, i.e., $r(i)$, will become $2^{r(i)} - 1$. Every operation of Algorithm 1 will occur on the transformed relevance vectors. It is very easy to see that every step in analysis of the algorithm will be valid by just considering the transformed relevance vectors to be some new relevance vectors with magnified relevance values. The fact that r was binary valued in *SumLoss* played no role in the analysis of the algorithm or Lemma 8. The properties which allowed the extension was that $g(\cdot)$ is composed of univariate, monotonic, scalar valued functions and $DCG(\sigma, r)$ is a linear function of $f(\sigma)$ and $g(r)$.

It is also interesting to note that $M(y) = \underset{\sigma}{\text{argmax}}\, f(\sigma) \cdot y = \underset{\sigma}{\text{argmin}}\, \sigma \cdot y$. Thus, no changes in the algorithm is required, other than simple transformation of relevance values.

14.1.3 Proof of Theorem 9

Following the proof of Theorem 7, modified for DCG, Eq.13 gives (for DCG):

$$\begin{aligned} E[\sum_{t=1}^{T} DCG(\sigma_t, r_t)] \geq \max_{\sigma} \sum_{t=1}^{T} DCG(\sigma, r_t) \\ - O(m^{1/3} R^{2/3} (DA)^{1/3} T^{2/3}). \end{aligned}$$

For DCG, $D = \sum_{i=1}^m f^s(\sigma(i)) = O(m), R = \sum_{i=1}^m f^s(\sigma(i)) g^s(r(i)) = O(m(2^n - 1)), A = \sum_{i=1}^m g^s(r(i)) = O(m(2^n - 1))$ and hence the regret is $O((2^n - 1)m^{5/3}T^{2/3})$.

14.2 Extension of Results of SumLoss to Prec@k

Since Prec@$k = f(\sigma) \cdot r$, with $f(\cdot)$ having properties enlisted in Sec. 3, all results of SumLoss trivially extend to Prec@k, except results on local observability. The reason is that while $f(\cdot)$ of SumLoss is strictly monotonic, $f(\cdot)$ of Prec@k is monotonic but not strict. The gain function depends only on the objects in the top-k position of the ranked list, irrespective of the order. A careful analysis shows that Lemma 3 fails to extend to the case of Prec@k. Thus, we cannot define the neighboring action set of the Pareto optimal action pairs, and hence cannot prove or disprove local observability. The structure of neighbors in Prec@k remains an open question.

However, the non-strict monotonicity of Prec@k is required for solving $M(y) = \underset{\sigma}{\operatorname{argmax}}\ f(\sigma) \cdot y$ efficiently.

14.2.1 Proof of Theorem 10

Following the proof of Theorem 7, modified for Prec@k, Eq.13 gives (for Prec@k):

$$E[\sum_{t=1}^{T} Prec@k(\sigma_t, r_t)] \geq \max_{\sigma} \sum_{t=1}^{T} Prec@k(\sigma, r_t) - O(m^{1/3} R^{2/3} (DA)^{1/3} T^{2/3}).$$

For Prec@k, $D = \sum_{i=1}^{k} f^s(\sigma(i)) = O(k), R = \sum_{i=1}^{m} f^s(\sigma(i)) g^s(r(i)) = O(k), A = \sum_{i=1}^{m} r(i) = O(m)$ and hence the regret is $O(km^{2/3}T^{2/3})$.

15 Non-existence of Sublinear Regret Bounds for NDCG, MAP and AUC

We show via simple calculations that for the case $m = 3$, global observability condition fails to hold for NDCG, when relevance vectors are restricted to take binary values. The intuition behind failure to satisfy global observability condition is that the $NDCG(\sigma, r) = f(\sigma) \cdot g(r)$, where where $g(r) = r/Z(r)$ (See Sec.3). Thus, $g(\cdot)$ cannot be by univariate, scalar valued functions. This makes it impossible to write the difference between two rows as linear combination of columns of (transposed) signal matrices.

15.1 Proof of Lemma 11

Proof. We will first consider NDCG and then, MAP and AUC.

NDCG:

The first and last row of Table 1, when calculated for NDCG, are:

$$\begin{aligned}\ell_{\sigma_1} =&[1, 1/2, 1/\log_2 3, (1 + \log_2 3/2))/(1 + \log_2 3), 1,\\ &3/(2(1 + 1/\log_2 3)), 1, 1]\\ \ell_{\sigma_6} =&[1, 1, \log_2 2/\log_2 3, 1, 1/2, 3/(2(1 + 1/\log_2 3)),\\ &(1 + (\log_2 3)/2))/(1 + \log_2 3), 1]\end{aligned}$$

We remind once again that NDCG is a gain function, as opposed to SumLoss.

The difference between the two vectors is:

$$\begin{aligned}\ell_{\sigma_1} - \ell_{\sigma_6} =&[0, -1/2, 0, -\log_2 3/(2(1 + \log_2 3)),\\ &1/2, 0, \log_2 3/(2(1 + \log_2 3)), 0].\end{aligned}$$

The signal matrices are same as SumLoss:

$$S_{\sigma_1} = S_{\sigma_2} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$S_{\sigma_3} = S_{\sigma_5} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$S_{\sigma_4} = S_{\sigma_6} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

It can now be easily checked that $\ell_{\sigma_1} - \ell_{\sigma_6}$ does not lie in the (combined) column span of the (transposed) signal matrices.

We show similar calculations for MAP and AUC.

MAP:

We once again take $m = 3$. The first and last row of Table 1, when calculated for MAP, is:

$$\begin{aligned}\ell_{\sigma_1} &= [1, 1/3, 1/2, 7/12, 1, 5/6, 1, 1]\\ \ell_{\sigma_6} &= [1, 1, 1/2, 1, 1/3, 5/6, 7/12, 1]\end{aligned}$$

Like NDCG, MAP is also a gain function.

The difference between the two vectors is:

$$\ell_{\sigma_1} - \ell_{\sigma_6} = [0, -2/3, 0, -5/12, 2/3, 0, 5/12, 0].$$

The signal matrices are same as in the SumLoss case:

$$S_{\sigma_1} = S_{\sigma_2} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$S_{\sigma_3} = S_{\sigma_5} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$S_{\sigma_4} = S_{\sigma_6} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

It can now be easily checked that $\ell_{\sigma_1} - \ell_{\sigma_6}$ does not lie in the (combined) column span of the (transposed) signal matrices. □

AUC:

For AUC, we will show the calculations for $m = 4$. This is because global observability does hold with $m = 3$, as the normalizing factors for all relevance vectors with non-trivial mixture of 0 and 1 are same (i.e, when relevance vector has 1 irrelevant and 2 relevant objects, and 1 relevant and 2 irrelevant objects, the normalizing factors are same). The normalizing factor changes from $m = 4$ onwards; hence global observability fails.

Table 1 will be extended since $m = 4$. Instead of illustrating the full table, we point out the important facts about the loss matrix table with $m = 4$ for AUC.

The 2^4 relevance vectors heading the columns are:

$r_1 = 0000,\ r_2 = 0001,\ r_3 = 0010,\ r_4 = 0100,\ r_5 = 1000,\ r_6 = 0011,\ r_7 = 0101,\ r_8 = 1001,\ r_9 = 0110,\ r_{10} = 1010,\ r_{11} = 1100,\ r_{12} = 0111,\ r_{13} = 1011,\ r_{14} = 1101,\ r_{15} = 1110,\ r_{16} = 1111.$

We will calculate the losses of 1st and last (24th) action, where $\sigma_1 = 1234$ and $\sigma_{24} = 4321$.

$$\ell_{\sigma_1} = [0, 1, 2/3, 1/3, 0, 1, 3/4, 1/2, 1/2, 1/4, 0, 1, 2/3, 1/3, 0, 0]$$
$$\ell_{\sigma_{24}} = [0, 0, 1/3, 2/3, 1, 0, 1/4, 1/2, 1/2, 3/4, 1, 0, 1/3, 2/3, 1, 0]$$

AUC, like SumLoss, is a loss function.

The difference between the two vectors is:

$$\ell_{\sigma_1} - \ell_{\sigma_{24}} = [0, 1, 1/3, -1/3, -1, 1, 1/2, 0, 0, -1/2, -1, 1, 1/3, -1/3, -1, 0].$$

The signal matrices for AUC with $m = 4$ will be slightly different. This is because there are 24 signal matrices, corresponding to 24 actions. However, groups of 6 actions will share the same signal matrix. For example, all 6 permutations that place object 1 first will have same signal matrix, all 6 permutations that place object 2 first will have same signal matrix, and so on. For simplicity, we denote the signal matrices as S_1, S_2, S_3, S_4, where S_i corresponds to signal matrix where object i is placed at top. We have:

$$S_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$S_4 = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

It can now be easily checked that $\ell_{\sigma_1} - \ell_{\sigma_{24}}$ does not lie in the (combined) column span of transposes of S_1, S_2, S_3, S_4.