Learning Policies for Contextual Submodular Prediction - Supplementary Material

Stephane Ross STEPHANEROSS@CMU.EDU
Jiaji Zhou JIAJIZ@ANDREW.CMU.EDU
Yisong Yue YISONGYUE@CMU.EDU
Debadeepta Dey DEBADEEP@CS.CMU.EDU
J. Andrew Bagnell DBAGNELL@RI.CMU.EDU
School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

A. Proofs of Theoretical Results

This appendix contains the proofs of the various theoretical results presented in this paper.

A.1. Preliminaries

We begin by proving a number of lemmas about monotone submodular functions, which will be useful to prove our main results.

Lemma 1. *Let $\mathcal{S}$ be a set and f be a monotone submodular function defined on list of items from $\mathcal{S}$. For any lists A, B, we have that:*

$$f(A \oplus B) - f(A) \leq |B|(\mathbb{E}_{s \sim U(B)}[f(A \oplus s)] - f(A))$$

for $U(B)$ the uniform distribution on items in B.

Proof. For any list A and B, let B_i denote the list of the first i items in B, and b_i the i^{th} item in B. We have that:

$$\begin{array}{ll} & f(A \oplus B) - f(A) \\ = & \sum_{i=1}^{|B|} f(A \oplus B_i) - f(A \oplus B_{i-1}) \\ \leq & \sum_{i=1}^{|B|} f(A \oplus b_i) - f(A) \\ = & |B|(\mathbb{E}_{b \sim U(B)}[f(A \oplus b)] - f(A)) \end{array}$$

where the inequality follows from the submodularity property of f. □

Lemma 2. *Let $\mathcal{S}$ be a set, and f a monotone submodular function defined on lists of items in $\mathcal{S}$. Let A, B be any lists of items from $\mathcal{S}$. Denote A_j the list of the first j items in A, $U(B)$ the uniform distribution on items in B and define $\epsilon_j = \mathbb{E}_{s \sim U(B)}[f(A_{j-1} \oplus s)] - f(A_j)$, the additive error term in competing with the average marginal benefits of the items in B when picking the j^{th} item in A (which could be positive or negative). Then:*

$$f(A) \geq (1-(1-1/|B|)^{|A|})f(B) - \sum_{i=1}^{|A|}(1-1/|B|)^{|A|-i}\epsilon_i$$

In particular if $|A| = |B| = k$, then:

$$f(A) \geq (1-1/e)f(B) - \sum_{i=1}^{k}(1-1/k)^{k-i}\epsilon_i$$

and for $\alpha = \exp(-|A|/|B|)$ (i.e. $|A| = |B|\log(1/\alpha)$):

$$f(A) \geq (1-\alpha)f(B) - \sum_{i=1}^{|A|}(1-1/|B|)^{|A|-i}\epsilon_i$$

Proof. Using the monotone property and previous lemma 1, we must have that: $f(B) - f(A) \leq f(A \oplus B) - f(A) \leq |B|(\mathbb{E}_{b \sim U(B)}[f(A \oplus b)] - f(A))$.

Now let $\Delta_j = f(B) - f(A_j)$. By the above we have that

$$\begin{array}{ll} & \Delta_j \\ \leq & |B|[\mathbb{E}_{s \sim U(B)}[f(A_j \oplus s)] - f(A_j)] \\ = & |B|[\mathbb{E}_{s \sim U(B)}[f(A_j \oplus s)] - f(A_{j+1}) \\ & +f(A_{j+1}) - f(B) + f(B) - f(A_j)] \\ = & |B|[\epsilon_{j+1} + \Delta_j - \Delta_{j+1}] \end{array}$$

Rearranging terms, this implies that $\Delta_{j+1} \leq (1 - 1/|B|)\Delta_j + \epsilon_{j+1}$. Recursively expanding this recurrence from $\Delta_{|A|}$, we obtain:

$$\Delta_{|A|} \leq (1-1/|B|)^{|A|}\Delta_0 + \sum_{i=1}^{|A|}(1-1/|B|)^{|A|-i}\epsilon_i$$

Using the definition of $\Delta_{|A|}$ and rearranging terms, we obtain $f(A) \geq (1-(1-1/|B|)^{|A|})f(B) - \sum_{i=1}^{|A|}(1 -$

$1/|B|)^{|A|-i}\epsilon_i$. This proves the first statement of the theorem. The following two statements follow from the observations that $(1-1/|B|)^{|A|} = \exp(|A|\log(1-1/|B|)) \leq \exp(-|A|/|B|) = \alpha$. Hence $(1-(1-1/|B|)^{|A|})f(B) \geq (1-\alpha)f(B)$. When $|A| = |B|$, $\alpha = 1/e$ and this proves the special case where $|A| = |B|$. □

For the greedy list construction strategy, the ϵ_j in the last lemma are always ≤ 0, such that Lemma 2 implies that if we construct a list of size k with greedy, it must achieve at least 63% of the value of the optimal list of size k, but also that it must achieve at least 95% of the value of the optimal list of size $\lfloor k/3 \rfloor$, and at least 99.9% of the value of the optimal list of size $\lfloor k/7 \rfloor$.

A more surprising fact that follows from the last lemma is that constructing a list stochastically, by sampling items from a particular fixed distribution, can provide the same guarantee as greedy:

Lemma 3. *Let $\mathcal{S}$ be a set, and f a monotone submodular function defined on lists of items in $\mathcal{S}$. Let B be any list of items from $\mathcal{S}$ and $U(B)$ the uniform distribution on elements in B. Suppose we construct the list A by sampling k items randomly from $U(B)$ (with replacement). Denote A_j the list obtained after j samples, and P_j the distribution over lists obtained after j samples. Then:*

$$\mathbb{E}_{A\sim P_k}[f(A)] \geq (1-(1-1/|B|)^k)f(B)$$

In particular, for $\alpha = \exp(-k/|B|)$:

$$\mathbb{E}_{A\sim P_k}[f(A)] \geq (1-\alpha)f(B)$$

Proof. The proof follows a similar proof to the previous lemma. Recall that by the monotone property and lemma 1, we have that for any list A: $f(B) - f(A) \leq f(A \oplus B) - f(A) \leq |B|(\mathbb{E}_{b\sim U(B)}[f(A \oplus b)] - f(A))$. Because this holds for all lists, we must also have that for any distribution P over lists A, $f(B) - \mathbb{E}_{A\sim P}[f(A)] \leq |B|\mathbb{E}_{A\sim P}[\mathbb{E}_{b\sim U(B)}[f(A \oplus b)] - f(A)]$. Also note that by the way we construct sets, we have that $\mathbb{E}_{A_{j+1}\sim P_{j+1}}[f(A_{j+1})] = \mathbb{E}_{A_j\sim P_j}[\mathbb{E}_{s\sim U(B)}[f(A_j \oplus s)]]$

Now let $\Delta_j = f(B) - \mathbb{E}_{A_j\sim P_j}[f(A_j)]$. By the above we have that:

$$\begin{aligned}
&\Delta_j \\
\leq\ & |B|\mathbb{E}_{A_j\sim P_j}[\mathbb{E}_{s\sim U(B)}[f(A_j \oplus s)] - f(A_j)] \\
=\ & |B|\mathbb{E}_{A_j\sim P_j}[\mathbb{E}_{s\sim U(B)}[f(A_j \oplus s)] - f(B) \\
& +f(B) - f(A_j)] \\
=\ & |B|(\mathbb{E}_{A_{j+1}\sim P_{j+1}}[f(A_{j+1})] - f(B) \\
& +f(B) - \mathbb{E}_{A_j\sim P_j}[f(A_j)]) \\
=\ & |B|[\Delta_j - \Delta_{j+1}]
\end{aligned}$$

Rearranging terms, this implies that $\Delta_{j+1} \leq (1-1/|B|)\Delta_j$. Recursively expanding this recurrence from Δ_k, we obtain:

$$\Delta_k \leq (1-1/|B|)^k \Delta_0$$

Using the definition of Δ_k and rearranging terms we obtain $E_{A\sim P_k}[f(A)] \geq (1-(1-1/|B|)^k)f(B)$. The second statement follows again from the fact that $(1-(1-1/|B|)^k)f(B) \geq (1-\alpha)f(B)$ □

Corollary 1. *There exists a distribution that when sampled k times to construct a list, achieves an approximation ratio of $(1-1/e)$ of the optimal list of size k in expectation. In particular, if A^* is an optimal list of size k, sampling k times from $U(A^*)$ achieves this approximation ratio. Additionally, for any $\alpha \in (0,1]$, sampling $\lceil k\log(1/\alpha)\rceil$ times must construct a list that achieves an approximation ratio of $(1-\alpha)$ in expectation.*

Proof. Follows from the last lemma using $B = A^*$. □

This surprising result can also be seen as a special case of a more general result proven in prior related work that analyzed randomized set selection strategies to optimize submodular functions (lemma 2.2 in (Feige et al., 2011)).

A.2. Proofs of Main Results

We now provide the proofs of the main results in this paper. We provide the proofs for the more general contextual case where we learn over a policy class $\tilde{\Pi}$. All the results for the context-free case can be seen as special cases of these results when $\Pi = \tilde{\Pi} = \{\pi_s | s \in \mathcal{S}\}$ and $\pi_s(x, L) = s$ for any state x and list L.

We refer the reader to the notation defined in section 3 and 5 for the definitions of the various terms used.

Theorem 2 . *Let $\alpha = \exp(-m/k)$ and $k' = \min(m,k)$. After T iterations, for any $\delta, \delta' \in (0,1)$, we have that with probability at least $1-\delta$:*

$$F(\overline{\pi}, m) \geq (1-\alpha)F(L^*_{\pi,k}) - \frac{R}{T} - 2\sqrt{\frac{2\ln(1/\delta)}{T}}$$

and similarly, with probability at least $1-\delta-\delta'$:

$$\begin{aligned}
F(\overline{\pi}, m) \geq\ & (1-\alpha)F(L^*_{\pi,k}) - \tfrac{\mathbb{E}[R]}{T} - \sqrt{\tfrac{2k'\ln(1/\delta')}{T}} \\
& -2\sqrt{\tfrac{2\ln(1/\delta)}{T}}
\end{aligned}$$

Proof.

$$\begin{aligned}
&F(\overline{\pi}, m)\\
=\ & \tfrac{1}{T}\textstyle\sum_{t=1}^{T} F(\pi_t, m)\\
=\ & \tfrac{1}{T}\textstyle\sum_{t=1}^{T} \mathbb{E}_{L_{\pi,m}\sim\pi_t}[\mathbb{E}_{x\sim D}[f_x(L_{\pi,m}(x))]]\\
=\ & (1-\alpha)\mathbb{E}_{x\sim D}[f_x(L^*_{\pi,k}(x))]\\
& -[(1-\alpha)\mathbb{E}_{x\sim D}[f_x(L^*_{\pi,k}(x))]\\
& -\tfrac{1}{T}\textstyle\sum_{t=1}^{T} \mathbb{E}_{L_{\pi,m}\sim\pi_t}[\mathbb{E}_{x\sim D}[f_x(L_{\pi,m}(x))]]]
\end{aligned}$$

Now consider the sampled states $\{x_t\}_{t=1}^T$ and the policies $\pi_{t,i}$ sampled i.i.d. from π_t to construct the lists $\{L_t\}_{t=1}^T$ and denote the random variables $X_t = (1-\alpha)(\mathbb{E}_{x\sim D}[f_x(L^*_{\pi,k}(x))] - f_{x_t}(L^*_{\pi,k}(x_t))) - \mathbb{E}_{L_{\pi,m}\sim\pi_t}[\mathbb{E}_{x\sim D}[f_x(L_{\pi,m}(x))]] - f_{x_t}(L_t)]$. If π_t is deterministic, then simply consider all $\pi_{t,i} = \pi_t$. Because the x_t are i.i.d. from D, and the distribution of policies used to construct L_t only depends on $\{x_\tau\}_{\tau=1}^{t-1}$ and $\{L_\tau\}_{\tau=1}^{t-1}$, then the X_t conditioned on $\{X_\tau\}_{\tau=1}^{t-1}$ have expectation 0, and because $f_x \in [0,1]$ for all state $x \in \mathcal{X}$, X_t can vary in a range $r \subseteq [-2,2]$. Thus the sequence of random variables $Y_t = \sum_{i=1}^{t} X_i$, for t =1 to T, forms a martingale where $|Y_t - Y_{t+1}| \leq 2$. By the Azuma-Hoeffding's inequality, we have that $P(Y_T/T \geq \epsilon) \leq \exp(-\epsilon^2 T/8)$. Hence for any $\delta \in (0,1)$, we have that with probability at least $1-\delta$, $Y_T/T \leq 2\sqrt{\frac{2\ln(1/\delta)}{T}}$. Hence we have that with probability at least $1-\delta$:

$$\begin{aligned}
&F(\overline{\pi}, m)\\
=\ & (1-\alpha)\mathbb{E}_{x\sim D}[f_x(L^*_{\pi,k}(x))]\\
& -[(1-\alpha)\mathbb{E}_{x\sim D}[f_x(L^*_{\pi,k}(x))]\\
& -\tfrac{1}{T}\textstyle\sum_{t=1}^{T} \mathbb{E}_{L_{\pi,m}\sim\pi_t}[\mathbb{E}_{x\sim D}[f_x(L_{\pi,m}(x))]]]\\
=\ & (1-\alpha)\mathbb{E}_{x\sim D}[f_x(L^*_{\pi,k}(x))]\\
& -[(1-\alpha)\tfrac{1}{T}\textstyle\sum_{t=1}^{T} f_{x_t}(L^*_{\pi,k}(x_t))\\
& -\tfrac{1}{T}\textstyle\sum_{t=1}^{T} f_{x_t}(L_t)] - Y_T/T\\
=\ & (1-\alpha)\mathbb{E}_{x\sim D}[f_x(L^*_{\pi,k}(x))]\\
& -[(1-\alpha)\tfrac{1}{T}\textstyle\sum_{t=1}^{T} f_{x_t}(L^*_{\pi,k}(x_t))\\
& -\tfrac{1}{T}\textstyle\sum_{t=1}^{T} f_{x_t}(L_t)] - 2\sqrt{\tfrac{2\ln(1/\delta)}{T}}
\end{aligned}$$

Let $w_i = (1-1/k)^{m-i}$. From Lemma 2, we have:

$$\begin{aligned}
& (1-\alpha)\tfrac{1}{T}\textstyle\sum_{t=1}^{T} f_{x_t}(L^*_{\pi,k}(x_t)) - \tfrac{1}{T}\sum_{t=1}^{T} f_{x_t}(L_t)\\
\leq\ & \tfrac{1}{T}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{m} w_i(\mathbb{E}_{\pi\sim U(L^*_{\pi,k})}[f_{x_t}(L_{t,i-1}\oplus\pi(x_t))]\\
& -f_{x_t}(L_{t,i}))\\
=\ & \mathbb{E}_{\pi\sim U(L^*_{\pi,k})}[\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{m} w_i(f_{x_t}(L_{t,i-1}\oplus\pi(x_t))\\
& -f_{x_t}(L_{t,i}))]\\
\leq\ & \max_{\pi\in\Pi}[\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{m} w_i(f_{x_t}(L_{t,i-1}\oplus\pi(x_t))\\
& -f_{x_t}(L_{t,i}))]\\
\leq\ & \max_{\pi\in\tilde{\Pi}}[\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{m} w_i(f(L_{t,i-1}\oplus\pi(x_t))\\
& -f_{x_t}(L_{t,i}))]\\
=\ & R/T
\end{aligned}$$

Hence combining with the previous result proves the first part of the theorem.

Additionally, for the sampled environments $\{x_t\}_{t=1}^T$ and the policies $\pi_{t,i}$, consider the random variables $Q_{m(t-1)+i} = w_i\mathbb{E}_{\pi\sim\pi_t}[f_{x_t}(L_{t,i-1}\oplus\pi(x_t, L_{t,i-1}))] - w_i f_{x_t}(L_{t,i})$. Because each draw of $\pi_{t,i}$ is i.i.d. from π_t, we have that again the sequence of random variables $Z_j = \sum_{i=1}^{j} Q_i$, for $j = 1$ to Tm forms a martingale and because each Q_i can take values in a range $[-w_j, w_j]$ for $j = 1 + \mathrm{mod}(i-1, m)$, we have $|Z_i - Z_{i-1}| \leq w_j$. Since $\sum_{i=1}^{Tm} |Z_i - Z_{i-1}|^2 \leq T\sum_{i=1}^{m}(1-1/k)^{2(m-i)} \leq T\min(k,m) = Tk'$, by Azuma-Hoeffding's inequality, we must have that $P(Z_{Tm} \geq \epsilon) \leq \exp(-\epsilon^2/2Tk')$. Thus for any $\delta' \in (0,1)$, with probability at least $1-\delta'$, $Z_{Tm} \leq \sqrt{2Tk'\ln(1/\delta)}$. Hence combining with the previous result, it must be the case that with probability at least $1-\delta-\delta'$, both $Y_T/T \leq 2\sqrt{\frac{2\ln(1/\delta)}{T}}$ and $Z_{Tm} \leq \sqrt{2Tk'\ln(1/\delta')}$ holds.

Now note that:

$$\begin{aligned}
& \max_{\pi\in\tilde{\Pi}}[\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{m} w_i(f(L_{t,i-1}\oplus\pi(x_t)) - f_{x_t}(L_{t,i}))]\\
=\ & \max_{\pi\in\tilde{\Pi}}[\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\sum_{i=1}^{m} w_i(f_{x_t}(L_{t,i-1}\oplus\pi(x_t))\\
& -\mathbb{E}_{\pi'\sim\pi_t}[f(L_{t,i-1}\oplus\pi'(x_t, L_{t,i-1}))])] + Z_{Tm}/T\\
=\ & \mathbb{E}[R]/T + Z_{Tm}/T
\end{aligned}$$

Using this additional fact, and combining with previous results we must have that with probability at least $1-\delta-\delta'$:

$$\begin{aligned}
& F(\overline{\pi}, m)\\
\geq\ & (1-\alpha)F(L^*_{\pi,k}) - [(1-\alpha)\tfrac{1}{T}\textstyle\sum_{t=1}^{T} f_{x_t}(L^*_{\pi,k}(x_t))\\
& -\tfrac{1}{T}\textstyle\sum_{t=1}^{T} f_{x_t}(L_t)] - 2\sqrt{\tfrac{2\ln(1/\delta)}{T}}\\
\geq\ & (1-\alpha)F(L^*_{\pi,k}) - \mathbb{E}[R]/T - Z_{Tm}/T - 2\sqrt{\tfrac{2\ln(1/\delta)}{T}}\\
\geq\ & (1-\alpha)F(L^*_{\pi,k}) - \mathbb{E}[R]/T - \sqrt{\tfrac{2k'\ln(1/\delta')}{T}}\\
& -2\sqrt{\tfrac{2\ln(1/\delta)}{T}}
\end{aligned}$$

□

We now show that the expected regret must grow with $\sqrt{k'}$ and not k', hen using Weighted Majority with the optimal learning rate (or with the doubling trick).

Corollary 2 . *Under the event where Theorem 2 holds (the event that occurs w.p. $1-\delta-\delta'$), if $\tilde{\Pi}$ is a finite set of policies, using Weighted Majority with the optimal learning rate guarantees that after T iterations:*

$$\begin{aligned}
\mathbb{E}[R]/T \quad \leq\ & \tfrac{4k'\ln|\tilde{\Pi}|}{T} + 2\sqrt{\tfrac{k'\ln|\tilde{\Pi}|}{T}}\\
& +2^{9/4}(k'/T)^{3/4}(\ln(1/\delta'))^{1/4}\sqrt{\ln|\tilde{\Pi}|}
\end{aligned}$$

For large enough T in $\Omega(k'(\ln|\tilde{\Pi}| + \ln(1/\delta')))$, we obtain that:

$$\mathbb{E}[R]/T \leq O(\sqrt{\frac{k' \ln|\tilde{\Pi}|}{T}})$$

Proof. We use a similar argument to Streeter & Golovin Lemma 4 (Streeter & Golovin, 2007) to bound $\mathbb{E}[R]$ in the result of theorem 2 . Consider the sum of the benefits accumulated by the learning algorithm at position i in the list, for $i \in 1, 2, \dots, m$, i.e. let $y_i = \sum_{t=1}^{T} b(\pi_{t,i}(x_t, L_{t,i-1})|x_t, L_{t,i-1})$, where $\pi_{t,i}$ corresponds to the particular sampled policy by Weighted Majority for choosing the item at position i, when constructing the list L_t for state x_t. Note that $\sum_{i=1}^{m}(1-1/k)^{m-i} y_i \leq \sum_{i=1}^{m} y_i \leq T$ by the fact that the monotone submodular function f_x is bounded in $[0,1]$ for all state x. Now consider the sum of the benefits you could have accumulated at position i, had you chosen the best fixed policy in hindsight to construct all list, keeping the policy fixed as the policy is constructed, i.e. let $z_i = \sum_{t=1}^{T} b(\pi^*(x_t, L_{t,i-1})|x_t, L_{t,i-1})$, for $\pi^* = \arg\max_{\pi \in \tilde{\Pi}} \sum_{i=1}^{m}(1 - 1/k)^{m-i} \sum_{t=1}^{T} b(\pi^*(x_t, L_{t,i-1})|x_t, L_{t,i-1})$ and let $r_i = z_i - y_i$. Now denote $Z = \sqrt{\sum_{i=1}^{m}(1-1/k)^{m-i} z_i}$. We have $Z^2 = \sum_{i=1}^{m}(1 - 1/k)^{m-i} z_i = \sum_{i=1}^{m}(1 - 1/k)^{m-i}(y_i + r_i) \leq T + R$, where R is the sample regret incurred by the learning algorithm. Under the event where theorem 2 holds (i.e. the event that occurs with probability at least $1 - \delta - \delta'$), we had already shown that $R \leq \mathbb{E}[R] + Z_{Tm}$, for $Z_{Tm} \leq \sqrt{2Tk' \ln(1/\delta')}$, in the second part of the proof of theorem 2 . Thus when theorem 2 holds, we have $Z^2 \leq T + \sqrt{2Tk' \ln(1/\delta')} + \mathbb{E}[R]$. Now using the generalized version of weighted majority with rewards (i.e. using directly the benefits as rewards) (Arora et al., 2012), since the rewards at each update are in $[0, k']$, we have that with the best learning rate in hindsight [1]: $\mathbb{E}[R] \leq 2Z\sqrt{k' \ln|\tilde{\Pi}|}$. Thus we obtain $Z^2 \leq T + \sqrt{2Tk' \ln(1/\delta')} + 2Z\sqrt{k' \ln|\tilde{\Pi}|}$. This is a quadratic inequality of the form $Z^2 - 2Z\sqrt{k' \ln|\tilde{\Pi}|} - T - \sqrt{2Tk' \ln(1/\delta')} \leq 0$, with the additional constraint $Z \geq 0$. This implies Z is less than or equal to the largest non-negative root of the polynomial $Z^2 - 2Z\sqrt{k' \ln|\tilde{\Pi}|} - T - \sqrt{2Tk' \ln(1/\delta')}$. Solving for the roots, we obtain

$$\begin{aligned} Z &\leq \sqrt{k' \ln|\tilde{\Pi}|} + \sqrt{k' \ln|\tilde{\Pi}| + T + \sqrt{2Tk' \ln(1/\delta')}} \\ &\leq 2\sqrt{k' \ln|\tilde{\Pi}|} + \sqrt{T} + (2Tk' \ln(1/\delta'))^{1/4} \end{aligned}$$

[1] if not a doubling trick can be used to get the same regret bound within a small constant factor (Cesa-Bianchi et al., 1997)

Plugging back Z into the expression $\mathbb{E}[R] \leq 2Z\sqrt{k' \ln|\tilde{\Pi}|}$, we obtain:

$$\begin{aligned} \mathbb{E}[R] \leq \quad & 4k' \ln|\tilde{\Pi}| + 2\sqrt{Tk' \ln|\tilde{\Pi}|} \\ & + 2(2T\ln(1/\delta'))^{1/4}(k')^{3/4}\sqrt{\ln|\tilde{\Pi}|} \end{aligned}$$

Thus the average regret:

$$\begin{aligned} \frac{\mathbb{E}[R]}{T} \leq \quad & \frac{4k' \ln|\tilde{\Pi}|}{T} + 2\sqrt{\frac{k' \ln|\tilde{\Pi}|}{T}} \\ & + 2^{9/4}(k'/T)^{3/4}(\ln(1/\delta'))^{1/4}\sqrt{\ln|\tilde{\Pi}|} \end{aligned}$$

For T in $\Omega(k'(\ln\tilde{\Pi} + \ln(1/\delta')))$, the dominant term is $2\sqrt{\frac{k' \ln|\tilde{\Pi}|}{T}}$, and thus $\frac{\mathbb{E}[R]}{T}$ is $O(\sqrt{\frac{k' \ln|\tilde{\Pi}|}{T}})$. □

Corollary 3 . *Let $\alpha = \exp(-m/k)$ and $k' = \min(m, k)$. If we run an online learning algorithm on the sequence of convex loss C_t instead of ℓ_t, then after T iterations, for any $\delta \in (0,1)$, we have that with probability at least $1 - \delta$:*

$$F(\overline{\pi}, m) \geq (1 - \alpha)F(L^*_{\pi,k}) - \frac{\tilde{R}}{T} - 2\sqrt{\frac{2\ln(1/\delta)}{T}} - \mathcal{G}$$

where $\tilde{R}$ is the regret on the sequence of convex loss C_t, and $\mathcal{G} = \frac{1}{T}[\sum_{t=1}^{T}(\ell_t(\pi_t) - C_t(\pi_t)) + \min_{\pi \in \tilde{\Pi}} \sum_{t=1}^{T} C_t(\pi) - \min_{\pi' \in \tilde{\Pi}} \sum_{t=1}^{T} \ell_t(\pi')]$ is the "convex optimization gap" that measures how close the surrogate losses C_t is to minimizing the cost-sensitive losses ℓ_t.

Proof. Follows immediately from Theorem 2 using the definition of R, $\tilde{R}$ and $\mathcal{G}$, since $\mathcal{G} = \frac{R - \tilde{R}}{T}$ □

References

Arora, S., Hazan, E., , and Kale, S. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 8:121–164, 2012.

Cesa-Bianchi, N., Freund, Y., Haussler, D., Helmbold, D. P., Schapire, R. E., and Warmuth, M. K. How to use expert advice. *Journal of the ACM*, 44(3): 427–485, May 1997.

Feige, U., Mirrokni, V. S., and Vondrak, J. Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, 40(4):1133–1153, 2011.

Streeter, Matthew and Golovin, Daniel. An online algorithm for maximizing submodular functions. Technical Report CMU-CS-07-171, Carnegie Mellon University, 2007.