CSCI 678: Theoretical Machine Learning
Homework 4

Fall 2024, Instructor: Haipeng Luo

This homework is due on **12/01, 11:59pm**. *See course website for more instructions on finishing and submitting your homework as well as the late policy. Total points: 50*

1. (**Stochastic MAB**) In Lecture 9, we proved that UCB achieves $\overline{\mathrm{Reg}}_n = \mathcal{O}(\sum_{a:\Delta_a>0} \frac{\ln n}{\Delta_a})$, which increases as the gaps decrease and make it harder to distinguish the optimal actions from the rest. However, if an action really has a tiny suboptimality gap, then by definition selecting it does not lead to large regret and thus there is really no point in distinguishing it from the optimal actions. Building on this intuition, in this problem you will further prove that UCB indeed also guarantees $\mathcal{O}(\sqrt{nK\ln n})$ pseudo regret at the same time, regardless of the values of the gaps.

   (a) (5pts) Recall (from Theorem 1 of Lecture 9) that with probability at least $1 - \frac{2K}{n}$, we have $m_n(a) \le \frac{16\ln n}{\Delta_a^2} + 1$ for every action a where $m_n(a)$ is the total number of times action a is pulled. Use this fact to prove the following pseudo regret bound

   $$\overline{\mathrm{Reg}}_n \le 3K + \Delta n + \sum_{a:\Delta_a>\Delta} \frac{16\ln n}{\Delta_a}, \tag{1}$$

   where $\Delta \in [0, 1]$ is an arbitrary threshold. (This matches the earlier intuition that we do not care about distinguishing an action with a small gap from the optimal action.)

   (b) (2pts) Pick an appropriate value of Δ and conclude the following bound

   $$\overline{\mathrm{Reg}}_n = \mathcal{O}(\sqrt{nK\ln n}).$$

   (You can assume $n \ge K$.)

2. (**Multiclass Perceptron**) In this exercise, you need to analyze variants of the Perceptron algorithm for *multiclass* classification, with either full information or bandit information. Specifically, consider a sequence of examples $x_1, \dots, x_n \in B_2^d$ with labels $y_1, \dots, y_n \in [K]$ where K is the number of possible classes. We assume that the following multiclass margin assumption holds: there exists a constant $\gamma > 0$ and K weight vectors $\theta_\star^1, \dots, \theta_\star^K \in B_2^d$ such that for each $t = 1, \dots, n$:

$$\left\langle \theta_\star^{y_t}, x_t \right\rangle \geq \left\langle \theta_\star^k, x_t \right\rangle + \gamma, \quad \forall k \neq y_t.$$

In other words, the predictor $\operatorname{argmax}_k \left\langle \theta^k, x_t \right\rangle$ makes perfect predictions for this dataset with γ margin. Now, consider the following learning protocol:

For $t = 1, \dots, n$:

- receive x_t and predict $s_t \in [K]$;
- observe $\begin{cases} y_t & \text{in the full-information setting} \\ \mathbf{1}\{s_t \neq y_t\} \text{ (i.e., if the prediction is correct)} & \text{in the bandit setting} \end{cases}$

In either case, we care about the total number of mistakes $M = \sum_{t=1}^n \mathbf{1}\{s_t \neq y_t\}$.

(a) In the full information setting, one can apply the following multiclass Perceptron algorithm, a natural generalization of its binary version studied in Lecture 7. Note that when the algorithm predicts correctly, the last update step in fact does nothing (similarly to the binary version).

Algorithm 1: Multiclass Perceptron

```
Initialize $\theta^1 = \cdots = \theta^K = \mathbf{0}$.
For $t = 1, \dots, n$:
    • receive $x_t$ and find $k_t \in \operatorname{argmax}_{k \in [K]} \langle \theta^k, x_t \rangle$;
    • predict $s_t = k_t$;
    • receive $y_t$ and update
          $\theta^{y_t} \leftarrow \theta^{y_t} + x_t$   and   $\theta^{k_t} \leftarrow \theta^{k_t} - x_t$.
```

Follow the steps below to prove $M \leq \frac{2K}{\gamma^2}$ for this algorithm.

i. (6pts) Similar to the binary case, we need to analyze the evolution of the quantities $\sum_{k=1}^K \left\langle \theta^k, \theta_\star^k \right\rangle$ and $\sum_{k=1}^K \left\| \theta^k \right\|_2^2$. To this end, denote the value of the weight vectors $\theta^1, \dots, \theta^K$ at the beginning of round t by $\theta_t^1, \dots, \theta_t^K$. Under the margin assumption, prove the following two facts for any $t = 1, \dots, n$:

$$\sum_{k=1}^K \left\langle \theta_{t+1}^k, \theta_\star^k \right\rangle \geq \sum_{k=1}^K \left\langle \theta_t^k, \theta_\star^k \right\rangle + \gamma \mathbf{1}\{s_t \neq y_t\},$$

and

$$\sum_{k=1}^K \left\| \theta_{t+1}^k \right\|_2^2 \leq \sum_{k=1}^K \left\| \theta_t^k \right\|_2^2 + 2\mathbf{1}\{s_t \neq y_t\}.$$

ii. (3pts) Combine the two facts in the last question to conclude $M \leq \frac{2K}{\gamma^2}$ (Hint: you will need to use the Cauchy-Schwarz inequality.)

(b) In the bandit setting, we make the following two changes to Algorithm 1: 1) first, in light of the exploration versus exploitation trade-off, it is natural to randomize the algorithm and explore every label with at least some small probability α; 2) second, the update $\theta^{y_t} \leftarrow \theta^{y_t} + x_t$ becomes invalid if the prediction is incorrect (since we do not know what y_t is), so we only do this update when we predict correctly, and we scale the update with the inverse probability of selecting the correct label, just like the idea of importance-weighted estimator in Exp3. The final algorithm is shown below.

```
Algorithm 2: Multiclass Perceptron with Bandit Feedback
Input: exploration parameter $\alpha \in (0, \frac{1}{2K}]$.
Initialize $\theta^1 = \cdots = \theta^K = \mathbf{0}$.
For $t = 1, \ldots, n$:
    • receive $x_t$ and find $k_t \in \mathrm{argmax}_{k \in [K]} \langle \theta^k, x_t \rangle$;
    • predict $s_t$ drawn from $p_t$, where $p_t(k) = (1 - \alpha K)\mathbf{1}\{k = k_t\} + \alpha, \forall k$;
    • receive $\mathbf{1}\{s_t \neq y_t\}$ and update
          $\theta^{y_t} \leftarrow \theta^{y_t} + \frac{x_t \mathbf{1}\{s_t = y_t\}}{p_t(y_t)}$   and   $\theta^{k_t} \leftarrow \theta^{k_t} - x_t$.
```

Follow the steps below to prove that this algorithm makes at most $\mathcal{O}\left(\frac{K\sqrt{n}}{\gamma^2}\right)$ mistakes in expectation. We will again use the notation $\theta_t^1, \ldots, \theta_t^K$ to denote the value of the weight vectors $\theta^1, \ldots, \theta^K$ at the beginning of round t, and study the evolution of the quantities $\sum_{k=1}^K \langle \theta^k, \theta_\star^k \rangle$ and $\sum_{k=1}^K \left\|\theta^k\right\|_2^2$ (*in expectation* this time).

i. (3pts) Under the margin assumption, prove the following for any $t = 1, \ldots, n$:

$$\mathbb{E}\left[\sum_{k=1}^K \langle \theta_{t+1}^k, \theta_\star^k \rangle\right] \geq \mathbb{E}\left[\sum_{k=1}^K \langle \theta_t^k, \theta_\star^k \rangle\right] + \gamma \mathbb{E}\left[\mathbf{1}\{k_t \neq y_t\}\right].$$

ii. (8pts) Next, prove the following for any $t = 1, \ldots, n$:

$$\mathbb{E}\left[\sum_{k=1}^K \left\|\theta_{t+1}^k\right\|_2^2\right] \leq \mathbb{E}\left[\sum_{k=1}^K \left\|\theta_t^k\right\|_2^2\right] + \frac{\mathbb{E}\left[\mathbf{1}\{k_t \neq y_t\}\right]}{\alpha} + 1.$$

Hint: consider the two cases $k_t \neq y_t$ and $k_t = y_t$ separately.

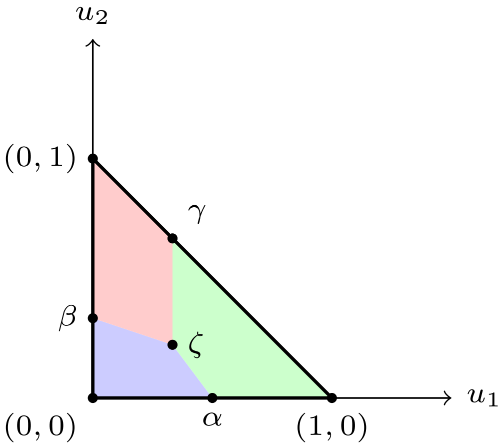
iii. (5pts) Combine the results from the last two questions to show $\gamma \mathbb{E}[N] \leq \sqrt{K\left(\frac{\mathbb{E}[N]}{\alpha} + n\right)}$ where $N = \sum_{t=1}^n \mathbf{1}\{k_t \neq y_t\}$. Further solve for $\mathbb{E}[N]$ to show $\mathbb{E}[N] \leq \frac{K}{\alpha\gamma^2} + \frac{\sqrt{Kn}}{\gamma}$.

iv. (4pts) Finally, use the result from the last step to prove $\mathbb{E}[M] \leq \frac{K}{\alpha\gamma^2} + \frac{\sqrt{Kn}}{\gamma} + \alpha n K$, and pick an appropriate value of α to conclude $\mathbb{E}[M] = \mathcal{O}\left(\frac{K\sqrt{n}}{\gamma} + \frac{K^2}{\gamma^2}\right)$.

3. (**Partial Monitoring**) Recall the dynamic pricing problem discussed in Lecture 9 and consider a simplified case with only 3 possible prices ($1, 2$, or 3 dollars). The loss matrix and feedback matrix are thus

$$\ell = \begin{pmatrix} 0 & 1 & 2 \\ c & 0 & 1 \\ c & c & 0 \end{pmatrix} \qquad \text{and} \qquad \Phi = \begin{pmatrix} \checkmark & \checkmark & \checkmark \\ \text{✗} & \checkmark & \checkmark \\ \text{✗} & \text{✗} & \checkmark \end{pmatrix}$$

for some storage cost $c > 0$. The cell decomposition of this problem is illustrated in the following picture, where we show the simplex $\Delta(3)$ by considering only the first two coordinates u_1 and u_2. Clearly, all 3 actions are Pareto-optimal, and every two actions are neighbors.



(a) (3pts) State which colored region in the cell decomposition picture corresponds to cell C_1, C_2, and C_3 respectively. Briefly explain why.

(b) (4pts) Calculate the coordinates of the four points α, β, γ, and ζ shown on the cell decomposition picture.

(c) (4pts) Prove that the following two action pairs are both locally observable: 1 and 2, 2 and 3.

(d) (3pts) The results from the last question imply that actions 1 and 3 must be globally observable. Now, prove that they are not locally observable. (This implies that this is a globally observable but not locally observable partial monitoring instance.)