AI3001: Advanced Topics in Machine Learning
Final Exam
October 25, 2022

Instructions:

- *The total number of marks is 25.*
- *The total duration of exam is 90 minutes. No cellphones or electronic aids are allowed. You can keep maximum one sheet of paper with formulas/notes.*

Problem 1. (FTL with Squared loss) Consider the problem of sequentially predicting a fixed and unknown sequence $y_1, y_2, \cdots$ in $\mathbb{R}^d$. At each round $t = 1, 2, \cdots, T$, the algorithm picks a point $p_t \in \mathbb{R}^d$ (knowing only $y_1, \cdots, y_{t-1}$). The current element y_t of the sequence is revealed, and the algorithm suffers a loss $f_t(p_t) = ||p_t - y_t||_2^2$.

Show that the FTL algorithm with squared loss function over a unit ball $B = \{x : ||x||_2 \leq 1\}$ [1] attains a logarithmic regret guarantee. Assume $y_0 = 0$ wherever needed. (5 marks)

Problem 2. Consider the following algorithm for K armed Bernoulli stochastic multi-armed bandits. The algorithm plays each arm once, and at each time $t > K$ an arm with highest lower confidence bound (LCB) on its mean reward is selected i.e.

$$i_t = \arg\max_i [\widehat{\mu}_{i,N_{i,t}}(t-1) - \sqrt{\frac{2\log(T)}{N_{i,t}}}]$$

here $N_{i,t}$ is the number of times arm i is pulled and $\widehat{\mu}_{i,N_{i,t}}(t-1)$ is the observed reward sample mean of arm i upto (and not including) time t. What kind of regret guarantee (in terms of the time horizon T) does this algorithm gets and why? (provide analytical proof to your claims) (4 marks)

Problem 3. (Two arm Bernoulli bandit with side information) Consider a stochastic 2-armed bandit problem where each arm i's reward sequence is generated independently from a Bernoulli distribution with parameter $\mu_i, i = 1, 2$. Further, it is known that $\mu_1 \neq \mu_2$ and $\mu_1, \mu_2 \in \{a, b\}$ where $0 < a < b < 1$ are known constants; the only uncertainty is in the order. Denote $\Delta := b - a$.

Consider the following (rather simple) algorithm. In the beginning, play each arm once, i.e., $i_1 = 1, i_2 = 2$. At every subsequent time $t \geq 3$, if there exists an arm whose observed empirical mean so far exceeds $(a + b)/2$, then play the arm with the highest empirical mean. Else, play both arms one after another, i.e., $i_t = 1$ followed by $i_{t+1} = 2$.[2]

1. Without loss of generality, let arm 1 be the optimal arm when running the algorithm. Split the set of times when arm 2 is played according to whether its observed empirical mean so far is (i) greater than or (ii) at most $(a + b)/2$. (2 marks)

---

[1] That is, both y_t and p_t are in B.

[2] If at time t the empirical rewards of both the arms are at-most $(a + b)/2$, skip the empirical rewards camparison at $t + 1$.

2. Bound (from above) the sum of probabilities of playing arm 2 at all times when event (i) occurs. (3 marks)

3. Bound (from above) the sum of probabilities of playing arm 2 at all times when event (ii) occurs by using the definition of the algorithm and relating event (ii) to an event involving the empirical mean of arm 1. (3 marks)

4. Put together the conclusions of the previous parts to derive a regret bound independent of T and depending only on Δ. (2 marks)

Problem 4. (Batch vs Online convex optimization) Consider a standard batch convex optimization problem: you want to find a minimum of the convex function $f : \mathbb{R} \to \mathbb{R}_+ \setminus \{\infty\}$ over a bounded convex set $\mathcal{K} \subset \mathbb{R}$ to within an accuracy $\varepsilon > 0$. In other words, you must output $x \in \mathcal{K}$ satisfying $f(x) \leq \min_{y \in \mathcal{K}} f(y) + \varepsilon$.

You are given an online algorithm ALG with the following property. For any number of rounds $t \geq 1$ and any sequence of non-negative loss functions $\{f_s\}_{s=1}^{t}$ from a family of convex functions $\mathcal{F}$, the algorithm's regret over [3] a convex set $\mathcal{K}$ is at t^α for some $\alpha \in [0, 1)$; i.e. ALG satisfies sublinear regret guarantee. How will you accomplish batch optimization objective using ALG? Show your work. (6 marks)

[3]The regret is computed with respect to any single point in $\mathcal{K}$.

Solution to Problem 1. NOte that FTL will predict $p_t = \frac{1}{t-1}\sum_{s=1}^{t-1} y_s$ and we have $p_{t+1} = \frac{1}{t}\sum_{s=1}^{t} y_s$. From FTL-BTL lemma we ahve

$$\mathcal{R}_T(FTL) \leq \sum_{t=1}^{T}[f_t(p_t) - f_t(p_{t+1})]$$

Observe

$$\begin{aligned}
f_t(p_t) - f_t(p_{t+1}) &= ||p_t - y_t||^2 - ||p_{t+1} - y_t||^2 \\
&= (p_t - p_{t+1}) \cdot (p_t + p_{t+1} - 2y_t) \\
&\leq 4||p_t - p_{t+1}|| \quad \text{(Cauchy-Shwartz inequality and the fact that } \mathcal{D} \text{ is a unit ball)}
\end{aligned}$$

$$\begin{aligned}
||p_t - p_{t+1}|| \leq ||\left(\frac{1}{t-1} - \frac{1}{t}\right)\sum_{s=1}^{t-1} y_s - \frac{y_t}{t}|| \\
&= \frac{1}{t}||\left(\frac{\sum_{s=1}^{t-1} y_s}{t-1} - \frac{y_t}{t}\right)|| \\
&\leq \frac{1}{t}\left(||\frac{\sum_{s=1}^{t-1} y_s}{t-1}|| + ||\frac{y_t}{t}||\right) \quad \text{(Triangle inequality)} \\
&\leq \frac{2}{t} \quad \text{(since } \mathcal{Y} \text{ is a unit ball)}
\end{aligned}$$

We obtained that $f_t(p_t) - f_t(p_{t+1}) \leq \frac{8}{t}$. This gives us logarithmic regret.

Recall that we applied the same technique for the follow the best expert with constant experts.

■

Solution to Problem 2. Consider K bandit problem with parameters $\mu_1 = 1/2$ and $\mu_i = 1/2 - \varepsilon$ for all $i \neq 1$. Observe that if any arm $i \neq 1$ obtains a reward 1 and arm 1 obtains reward 0, in the first pull of respective arms then arm 1 will never be pulled in future. This is true because after one pull the empirical rewards become 1 and 0 for arms i and 1 respectively and the second term in the LCB is negative and hence always increases with increasing $N_{i,t}$. Thus, irrespective of the rewards obtained for arm i in future pulls, arm 1 will never be pulled. Let us compute the probability of this event,

$$
\begin{aligned}
\mathcal{R}_T(LCB) &\geq \mathbb{P}(\text{arm 1 receives reward 0 and any arm} i \neq 1 \text{receives reward 1}) \\
&= 1/2 \cdot (1 - (1/2 + \varepsilon)^{K-1}) \cdot T \\
&\approx T/2 \quad \text{for } K \text{ large and } \varepsilon \text{ small}
\end{aligned}
$$

Hence, the worst case expected regret of LCB can be linear. ■

Solution of Problem 3.

1. Without loss of generality, let arm 1 be the optimal arm. Split the set of times when arm 2 (suboptimal arm) is played according to its observed empirical mean so far as follows:

   (1)

   $$\{i_t = 2\} = \{2\} \cup \underbrace{\{\ \widehat{\mu}_{2,N_{2,t}} > (a+b)/2, i_t = 2, t \geq 3\}}_{E1} \cup \underbrace{\{\widehat{\mu}_{2,N_{2,t}} \leq (a+b)/2, i_t = 2, t \geq 3\}}_{E2}$$

2. Consider E_1 and observe

$$
\begin{aligned}
\sum_{t=1}^{T} \mathbb{P}(\widehat{\mu}_{2,N_{2,t}} > (a+b)/2, i_t = 2, t \geq 3) &\leq \sum_{t=1}^{T} \mathbb{P}(\widehat{\mu}_{2,t} > (a+b)/2) \quad \text{(since } N_{i,1} \leq t\text{)} \\
&= \mathbb{P}(\widehat{\mu}_{2,t} - a > \Delta/2) \quad \text{(subtracting both sides by } a \text{ and definition of } \Delta\text{)} \\
&\leq \sum_{n=1}^{T} exp(-2n\Delta^2/4) \quad \text{(Hoeffding's Inequality)} \\
&= \frac{1}{exp(\Delta^2/2) - 1} \quad \text{(upper bound by using infinite sum)} \\
&< 2/\Delta^2 \quad \text{(Using } e^x \geq 1 + x\text{)}
\end{aligned}
$$

3. Consider $E2$

$$
\begin{aligned}
\sum_{t=1}^{T} \mathbb{P}(\widehat{\mu}_{2,N_{2,t}} \leq (a+b)/2, i_t = 2, t \geq 3) &\leq \sum_{t=3}^{T} (\widehat{\mu}_{1,N_{1,t-1}} \leq (a+b)/2, i_{t-1} = 1) \\
&\leq \sum_{n=1}^{T} \mathbb{P}(\widehat{\mu}_{1,n} \leq (a+b)/2) \\
&\leq \sum_{n=1}^{T} exp(-2n\Delta^2/4) \\
&= \frac{1}{exp(\Delta^2/2) - 1} \\
&< 2/\Delta^2
\end{aligned}
$$

4. Putting everything together

$$
\begin{aligned}
\mathcal{R}_T(\textsc{Alg}) &= \sum_{t=1}^{T} \Delta \cdot \mathbb{P}(i_t = 2) \\
&= \Delta + \sum_{t=3}^{T} \Delta \cdot \mathbb{P}(\widehat{\mu}_{2,N_{2,t}} > \frac{a+b}{2}, i_t = 2) + \sum_{t=3}^{T} \Delta \cdot \mathbb{P}(\widehat{\mu}_{2,N_{2,t}} \leq \frac{a+b}{2}, i_t = 2) \\
&\leq \Delta + \frac{4}{\Delta}
\end{aligned}
$$

■

Solution of Problem 4.

We know from the regret guarantee of ALG that the below equation holds for every $t \geq 1$

$$\sum_{s=1}^{t} f_s(w_s) - \sum_{s=1}^{t} f_s(w^*) \leq t^{\alpha}$$

Here $w^* = \arg\min_{w \in \mathcal{K}} \sum_{s=1}^{t} f_s(w^*)$.

Now feed the convex optimization algorithm ALG, $f_s = f$ for all s and observe the sequence of choices of ALG, $\{w_s\}_{s\geq 1}$. The regret guarantee then becomes

$$\sum_{s=1}^{t} f(w_s) - t \cdot f(w^*) \leq t^\alpha$$

Let $w^t := \frac{1}{t}\sum_{s=1}^{t} w_s$ and notice that since $\mathcal{K}$ is convex, $w^t \in \mathcal{K}$. Furthermore we have from Jensens inequality that $f(w^t) = f(\frac{1}{t}\sum_{s=1}^{t} w_s) \leq \frac{1}{t}\sum_{s=1}^{t} f(w_s)$. Hence for any t

$$f(w^t) \leq f(w^*) + t^{\alpha-1}$$

Notice that the above inequality holds for any $t \geq 1$. Set the stopping time $T = \lceil (1/\varepsilon)^{1-\alpha} \rceil$ and output the solution $x = w^T = \frac{1}{T}\sum_{s=1}^{T} w_s$. ■