AI4010: Online Learning
First Midterm Exam
Aug 2025

Instructions:

- *The total number of marks is 20.*

- *The total duration of the exam is 90 minutes. No electronic aids are allowed. You can keep a maximum of one sheet of paper with formulas/notes.*

- *All questions are mandatory. A yes/no answer without proper proof or justification will be given zero marks even if it is correct.*

- *Any plagiarism case, if detected, will attract F grade in the course irrespective of overall performance.*

Problem 0.1 (3 marks). *Show that a straightforward extension of the* MAJORITY *algorithm makes* $O(m \log(N))$ *mistakes when the best expert makes* $m > 0$ *mistakes.*

Solution. Run the algorithm in m phases. In each phase, starting with all the experts, eliminate the experts based on Majority rule. Go to the next phase when all the experts have made a mistake. After, m phases all the experts would have made m mistakes. Run the majority algorithm for one last time till you get the active experts to be 1. In total we ahve $m = 1$ phases and algorithm makes at-most $\log(T)$ mistakes in each phase, giving the desired result. ∎

Problem 0.2 (3 marks). *Let* $f(x) = e^{-\eta \ell(x,y)}$ *and consider the loss function* $\ell(x, y) = ||x - y||_2^2$ *over the closed ball of radius* $r > 0$ *and centered at origin i.e.* $B_r(0) \subseteq \mathbb{R}$. *Find the range of values of* η *for which* f *is concave.*

Solution. Let $d = 1$, use the fact that $f(x) := e^{-\eta ||x-y||_2^2}$ is twice differentiable. It is enough to to show that $\nabla^2 f(x) \leq 0$. Differentiate f twice to obtain $\nabla^2 f(x) = 2\eta e^{-2\eta(x-y)^2}(2\eta(x - y)^2 - 1)$. For this value to be ≤ 0 for all $x, y \in B_r(0)$ we need $\eta \leq \frac{1}{2(x-y)^2}$. IN the worst case when x and

y are diametrically opposite we have $(x - y)^2 = 4r^2$ which means $\eta \le \frac{1}{8r^2}$. For $d > 1$ is a easy extension where we need to show that $y^T \nabla^2 f(x) y \le 0$ for every y. ∎

Problem 0.3 (6 marks). *Prove or disprove the below statements.*

1. *LogSumExp (LSE) function is convex. [2 marks][1]*

2. *Let the function f be α_1-strongly convex and g be α_2-strongly convex then $f+g$ is $\alpha_1 + \alpha_2$-strongly convex. [2 marks]*

3. *If the function is α-exp-concave for some $\alpha > 0$ then it is also α'-exp-concave for any $\alpha' \in (0, \alpha]$. [2 marks]*

Solution.

1 True. To prove convexity observe

$$\nabla^2 f(x) = \frac{1}{1^T z}\left(\text{diag}(z) - \frac{zz^T}{1^T z}\right)$$

Here z is a vector with $z_i = e^{x_i}$. To prove convexity, it is enough to show $y^T \nabla^2 f(x) y \ge 0$ for any y. Towards this note that $y^T \nabla^2 f(x) y = \frac{(\sum_i z_i y_i^2)(\sum_i z_i) - (\sum_i y_i z_i)^2}{(\sum_i y_i)^2}$. Obtain the result using Cauchy-Shwartz inequality.

2 True. Easy to see from the definition.

3 True.

$$e^{-\alpha f(\theta x + (1-\theta)y)} \ge \theta e^{-\alpha f(x)} + (1 - \theta)e^{-\alpha f(y)} \quad \text{(exp concavity of } f)$$

$$\implies -\alpha(\theta f(x) + (1 - \theta)f(y)) \ge \log(\theta e^{-\alpha f(x)} + (1 - \theta)e^{-\alpha f(y)})$$

$$\ge \theta \log(e^{-\alpha f(x)}) + (1 - \theta) \log(e^{-\alpha f(y)})$$

$$= -\alpha(\theta f(x) + (1 - \theta)f(y))$$

---

[1]Log-sum-exp is defined as
$$LSE(a_1, a_2, \cdots a_n) = \log\left(\sum_{i=1}^{n} e^{a_i}\right)$$

Problem 0.4 (8 marks). *We developed two important tools in this course; randomization and doubling trick. In this question, we will show the effectiveness of these tools with an example.*

Consider that there is a Cow located at $s = 0$ on one side of an infinite fence (infinite towards both left and right ends). There is a hole on the fence at some location $t \in \mathbb{R}$ and the Cow wants to go to the other side of the fence as the Cow believes that the grass is greener on the other side (as always!!). Cow's goal is to find the hole by travelling least possible distance. Had the Cow known t in advance, she would have found the hole by travelling $|t|$ distance in the direction of the hole. Let $\text{OPT} := |t|$. We are interested in minimizing the competitive ratio defined as $CR_{Alg} := \frac{\text{ALG}}{\text{OPT}}$ where ALG is the distance travelled by the Cow with algorithm $\mathbf{Alg}$ in the absence of knowledge about location of the hole.

*To begin, consider the following **Bad** algorithm. Cow turns towards right (left) and keeps walking in hope of finding the hole. It is straightforward to see that the worst case competitive ratio of this strategy/algorithm is ∞ as the cow will never find the hole if it is on the left (right). It is clear that if the Cow has to find the hole it must explore to both right and left sides. Suppose now Cow decides to do the following;*

```
d₁ distance right → origin → d₂ distance left → origin → d₃ distance
right → ...
```

Answer the following questions,

1. *(2 marks) Let $d_k = k$ for all $k \geq 1$. Call this algorithm **Linear**. Show that the worst case competitive ratio CR_{Linear} can be arbitrarily large; i.e., for any finite $m > 0$, there exists an instance (location t) such that $CR_{Linear} > m$.*

2. *(3 marks) Let $d_1 = 1$ and $d_k = 2d_{k-1}$ for all $k > 1$. Call this algorithm **Doubling**. What is worst case $CR_{Doubling}$?*

3. *(3 marks) Cow tosses a fair Coin to decide between right-first doubling strategy (as in above question) and left-first doubling strategy (complementary strategy where cow starts with moving towards left instead of right). Call this algorithm $\mathbf{R + D}$. What is worst case expected CR_{R+D}?*

   Observe that both doubling trick and randomization help in improving the expected distance traveled to find the hole.

Solution.

1. Fix any integer $k \geq 1$ and consider placing the hole at $t = k$. Assume without loss of generality that the hole is on the right. Under the linear sequence $d_i = i$ the cow's rightward excursion lengths are the odd indices $d_1 = 1, d_3 = 3, d_5 = 5, \ldots$.

   The smallest odd index k_{odd} with $d_{k_{odd}} \geq t$ can be chosen as $k_{odd} = k$ by taking k odd.

The cow travels the full round-trip distance $2d_i$ for every excursion prior to the successful one, and in the successful excursion it travels exactly t (it stops when it hits the hole). Thus the total distance travelled is

$$\text{ALG} = \sum_{i=1}^{k-1} 2d_i + t.$$

With $d_i = i$ and $t = k$ (and choosing the indices so $d_k = k$ is the successful excursion) we compute

$$\text{ALG} = 2\sum_{i=1}^{k-1} i + k = 2 \cdot \frac{(k-1)k}{2} + k = k^2.$$

Therefore

$$CR_{\text{Linear}} = \frac{\text{ALG}}{\text{OPT}} = \frac{k^2}{k} = k.$$

Since k can be taken arbitrarily large, the worst-case competitive ratio of the linear strategy is unbounded.

2. Consider the deterministic doubling strategy that alternates sides starting, say, to the right. Fix a target $t > 0$ (right side). Let k be the index of the first *rightward* excursion whose nominal length $d_k = 2^{k-1}$ is at least t. Because rightward excursions occur on every odd index when starting right-first, k will be odd; but we only need the inequalities below.

The cow performs all excursions $1, 2, \ldots, k-1$ fully (each such excursion except the last successful one is a round-trip of length $2d_i$), and on the k-th excursion it travels distance exactly t (not necessarily the full d_k) and stops. Hence

$$\text{ALG} = \sum_{i=1}^{k-1} 2d_i + t = 2\sum_{i=1}^{k-1} 2^{i-1} + t = 2(2^{k-1} - 1) + t = 2^k - 2 + t.$$

We next bound $\text{OPT} = t$ from below using the previous rightward excursion length. The previous rightward excursion (the previous time we went right) had length $d_{k-2} = 2^{k-3}$ (this index exists because k is odd and $k \geq 3$ for nontrivial cases). Since k is the first right excursion with $d_k \geq t$, we must have

$$t > d_{k-2} = 2^{k-3}.$$

Therefore

$$CR_{\text{Doubling}}(t) = \frac{\text{ALG}}{t} = \frac{2^k - 2 + t}{t} = \frac{2^k - 2}{t} + 1 \le \frac{2^k}{2^{k-3}} + 1 = 8 + 1 = 9,$$

3. Let the randomized algorithm choose right-first or left-first each with probability 1/2. Fix a target $t > 0$ WLOG. Let k be the index as in part (2) defined relative to the right-first ordering: the first rightward excursion whose nominal length $d_k = 2^{k-1}$ satisfies $d_k \ge t$. As in part (2) we have $t \in (2^{k-3}, 2^{k-1}]$ and the right-first cost is

$$C_{\text{right-first}} = 2^k - 2 + t.$$

We now compute the cost when the algorithm *starts* to the left (left-first). There are two subcases depending on t:

- If $t \in (2^{k-3}, 2^{k-2}]$, then under left-first the rightward excursion that reaches t is the one with nominal length $d_{k-1} = 2^{k-2}$ (it occurs *earlier* in index compared to the right-first ordering). In this subcase

$$C_{\text{left-first}} = 2\sum_{i=1}^{k-2} d_i + t = 2(2^{k-2} - 1) + t = 2^{k-1} - 2 + t.$$

- If $t \in (2^{k-2}, 2^{k-1}]$, then under left-first the rightward excursion that finally reaches t is the later one (index $k+1$ in the right-first indexing), and

$$C_{\text{left-first}} = 2\sum_{i=1}^{k} d_i + t = 2(2^k - 1) + t = 2^{k+1} - 2 + t.$$

The adversary (who chooses t knowing the algorithm but not the coin flip) will pick t to maximize the expected cost. The expected cost for a given t equals

$$\mathbb{E}[\text{ALG}] = \frac{1}{2}C_{\text{right-first}} + \frac{1}{2}C_{\text{left-first}}.$$

We examine the worst-case over the two subcases.

Case A: $t \in (2^{k-3}, 2^{k-2}]$. Here

$$\mathbb{E}[\text{ALG}] = \frac{1}{2}(2^k - 2 + t) + \frac{1}{2}(2^{k-1} - 2 + t) = \frac{3 \cdot 2^{k-1} - 4 + 2t}{2}.$$

Thus the expected competitive ratio is

$$\frac{\mathbb{E}[\text{ALG}]}{t} = \frac{3 \cdot 2^{k-1} - 4 + 2t}{2t}.$$

This expression is decreasing in t, so it is maximized when t is as small as possible, i.e. $t \downarrow 2^{k-3}$. Substituting $t = 2^{k-3}$ (limit) gives

$$\lim_{t \downarrow 2^{k-3}} \frac{\mathbb{E}[\text{ALG}]}{t} = \frac{3 \cdot 2^{k-1} - 4 + 2 \cdot 2^{k-3}}{2 \cdot 2^{k-3}} = \frac{3 \cdot 2^{k-1} + 2^{k-2}}{2^{k-2}} = 3 \cdot 2 + 1 = 7.$$

Case B: $t \in (2^{k-2}, 2^{k-1}]$. Here

$$\mathbb{E}[\text{ALG}] = \frac{1}{2}(2^k - 2 + t) + \frac{1}{2}(2^{k+1} - 2 + t) = \frac{3 \cdot 2^k - 4 + 2t}{2},$$

so

$$\frac{\mathbb{E}[\text{ALG}]}{t} = \frac{3 \cdot 2^k - 4 + 2t}{2t}.$$

Again this is decreasing in t, so the worst value in this regime is achieved at the left endpoint $t \downarrow 2^{k-2}$. Substituting $t = 2^{k-2}$ (limit) yields

$$\lim_{t \downarrow 2^{k-2}} \frac{\mathbb{E}[\text{ALG}]}{t} = \frac{3 \cdot 2^k - 4 + 2 \cdot 2^{k-2}}{2 \cdot 2^{k-2}} = \frac{3 \cdot 2^k + 2^{k-1}}{2^{k-1}} = 3 \cdot 2 + 1 = 7.$$

∎