Problem Set 2

Note: We will discuss the first problem in the problem-solving session. However, you still need to write your own solution to every problem.

Problem 1: The doubling trick .... (10 points)

When running the exponential weight (EW) algorithm, the choice of ε requires the knowledge of n. However, we can use a standard *doubling trick* to eliminate this requirement.

Suppose we have an EW algorithm for an online learning problem with payoff in $[0, h]$ that has the regret bound analyzed in class. Without the knowledge of n, we run the following algorithm:

> For $m = 0, 1, 2, \ldots$: Instantiate an EW algorithm with time horizon 2^m and run it for 2^m rounds, until the learning process is ended.

Show that there exists an absolute constant c such that the (per-round) regret of the new algorithm is at most

$$ch\sqrt{(\log k)/n}.$$

Explain your answer. (You can aim for $c = 10$.)

Hint: Recall that the per-round regret of the EW algorithm is $2h\sqrt{(\log k)/n}$. You might want to use the *sub-additivity* of $\max(\cdot)$: For any sequence $(X_i)_{i=1}^n, (Y_i)_{i=1}^n$ we have $\max_i X_i + \max_i Y_i \geq \max_i(X_i + Y_i)$. You might also need to compute a geometric sum: $\sum_{k=0}^n ar^k = \frac{a(1-r^{n+1})}{1-r}$.

Problem 2: MAB with b-element sets as actions .... (10 points)

Consider the following multi-armed bandit problem. There are k actions $j = 1, 2, \ldots, k$. In each round $i = 1, 2, \ldots, n$, the learner chooses a subset S^i of actions of size b. The payoff of action j in round i is v_j^i. After the learner chooses the set S^i, she receives the total payoff of the chosen actions, $\sum_{j \in S^i} v_j^i$, and the individual payoffs of only the chosen actions $\{v_j^i : j \in S^i\}$ are revealed.

The learner's regret is defined as

$$\max_{S \subseteq \{1,2,\ldots,k\}, |S|=b} \sum_{i=1}^n \sum_{j \in S} v_j^i - \mathop{\mathbb{E}}_{S^1,\ldots,S^n} \left[\sum_{i=1}^n \sum_{j \in S^i} v_j^i \right].$$

Design a MAB algorithm such that the regret is at most

$$chb(k \log k)^{1/3} n^{2/3},$$

where c is an absolute constant of your choice. In particular, the regret must have a polynomial dependence on b. Explain your answer.