

COMS 6998-4 F17 Homework 1 (due Wednesday October 18)

Daniel Hsu (djh2164)

Instructions

Solve **two** of the assigned problems.

Submit your write-up on Gradescope as a PDF document by 11:59 PM of the due date. Upon submitting to Gradescope, make sure each problem has some page of the PDF assigned to it, or we will assume the problem was not solved.

Make sure the following appears on the first page of your write-up:

- your name,
- your UNI, and
- the names and UNIs of any students with whom you discussed the assignment.

You are welcome to use the Markdown or LaTeX source for the assignment as a template for your write-up. I use Pandoc <http://pandoc.org> to translate the Markdown to \LaTeX and ultimately to PDF.

Problem 1

Consider the setting for online decision-making with bandit feedback and expert advice, with the following change (parameterized by a positive integer n that is at most N). In each round, the learner chooses at most n of the experts, and receives the actions of only the chosen experts; the actions of the experts not chosen are unobserved. The rest of the round is the same as before: the learner choose an action and then receives the cost for (only) the chosen action. Assume the costs for all K actions are always in the range $[0, 1]$.

Prove that there is an algorithm for the learner in this setting with expected regret at most

$$O\left(\sqrt{\frac{\min\{K, n\}N \log N}{n} \cdot T}\right)$$

after T rounds.

Hint: Modify the Exp4 algorithm.

Your solution:

Problem 2

Let $(X_t)_{t=1,2,\dots}$ be a sequence of random variables, and $(Y_t)_{t=1,2,\dots}$ a sequence of random functionals such that, for each t , Y_t depends only on X_1, \dots, X_t . Define the notation $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot \mid X_1, \dots, X_{t-1}]$.

- Part 1. Prove that for any $T \geq 1$, $\lambda > 0$, and $\delta \in (0, 1)$,

$$\mathbb{P} \left(\sum_{t=1}^T Y_t \geq \frac{1}{\lambda} \sum_{t=1}^T \ln \mathbb{E}_t[\exp(\lambda Y_t)] + \frac{1}{\lambda} \ln(1/\delta) \right) \leq \delta.$$

Hint: Determine the expectation of the random variable

$$\exp \left(\lambda \sum_{t=1}^T Y_t - \sum_{t=1}^T \ln \mathbb{E}_t[\exp(\lambda Y_t)] \right),$$

and then apply Markov's inequality.

- Part 2 (optional). Assume that for all t , we have $Y_t \leq 1$ and $\mathbb{E}_t[Y_t] = 0$. Use the inequality from Part 1 to prove that for any $\delta \in (0, 1)$ and $v > 0$,

$$\mathbb{P} \left(\sum_{t=1}^T Y_t \geq \sqrt{2v \ln(1/\delta)} + \frac{\ln(1/\delta)}{3} \wedge \sum_{t=1}^T \mathbb{E}_t[Y_t^2] \leq v \right) \leq \delta.$$

Hint: First prove that for any $\lambda > 0$,

$$\frac{1}{\lambda} \sum_{t=1}^T \ln \mathbb{E}_t[\exp(\lambda Y_t)] \leq \frac{e^\lambda - \lambda - 1}{\lambda} \sum_{t=1}^T \mathbb{E}_t[Y_t^2].$$

Now use the fact that for $\lambda \in (0, 3)$,

$$\frac{e^\lambda - \lambda - 1}{\lambda} \leq \frac{\lambda}{2(1 - \lambda/3)}.$$

(You may use this last inequality without proof.) Conclude the proof by assuming $\sum_{t=1}^T \mathbb{E}_t[Y_t^2] \leq v$ and choosing $\lambda \in (0, 3)$ appropriately.

- Part 3. Consider a variant of the Exp4 algorithm for online decision-making with bandit feedback and expert advice in which all of the probability vectors p_t are chosen so that $p_t(a) \geq p_{\min}$ for some given $p_{\min} > 0$. (Don't worry about how this is done.) For each $a \in [K]$ and $t = 1, \dots, T$, assume that $c_t(a) \in [0, 1]$ and also that $\hat{c}_t(a)$ is the Horvitz-Thompson (inverse probability weighting) estimator of $c_t(a)$. Use either the probability inequality from Part 2 or Theorem 1.6 from David Freedman's 1975 paper "On Tail Probabilities for Martingales" to prove a bound on

$$\max_{i \in [N]} \left\{ \left(\sum_{t=1}^T c_t(a_t) - c_t(b_{t,i}) \right) - \left(\sum_{t=1}^T \langle \hat{c}_t, p_t \rangle - \hat{c}_t(b_{t,i}) \right) \right\}$$

that holds with probability at least $1 - \delta$.

Your solution:

Problem 3

Design an online classification algorithm such that, in the realizable setting where the target concept is a homogeneous linear threshold function, $c = h_{w^*}$ for some $w^* \in \mathbb{R}^d$, and the conditions below hold, a learner using the algorithm makes at most $\log_2[(1 + \|w^*\|_2)^d]$ mistakes.

The conditions are:

$$\|x_n\|_2 = 1, \quad |\langle w^*, x_n \rangle| \geq 1, \quad n = 1, 2, \dots$$

Your solution: