Fall 2022 *Instructor: Jacob Abernethy, Bhuvesh Kumar, Guanghui Wang*

ECE 8803: Online Decision Making in Machine Learning

Homework 3

Problem 1 (Boosting via Game Playing) 20 points

(a) (5pts):

- 0pts: if no meaningful work is given.
- 2pts: if the main ideas are correct.
- 4pts: if the construction of M is correct, but but made a small mistake in the final steps.
- 5pts: for fully correct answer.

(b) (5pts): The answer is unique.

(c) (5pts): Same as (a).

(d) (5pts): A correct application of the minimax theorem suffices for full credit.

Problem 2 (Understanding the pseudo-regret of greedy, pure-exploration and explore-then-commit) 20 points

(a) (4pts) Any reasoning that evaluates the probability as (at least) 0.06 receives full credit.

(b) (4pts) Either of the formulas for pseudo-regret can be used to get full credit. (The formula in terms of the expected number of times a suboptimal arm is pulled is the easier one, though.)

(c) (4pts) Either of the formulas for pseudo-regret can be used to get full credit. (The formula in terms of the expected number of times a suboptimal arm is pulled is the easier one, though.)

(d) (4pts) The correct application of Hoeffding's bound gets full credit for this sub-part. The constants in the bound do not need to exactly match the solution (i.e. they can be higher than the provided solution), as long as they are rigorously argued, for full credit.

(e) (4pts) See part (d).

(f) Correct repetition of parts (a) – (e)) gets 5 bonus points. You get 1 bonus point each for each correct repetition (which involves the same correct reasoning as the corresponding sub-part together with the correct dependence on $\Delta = p_2 - p_1$). The second part of the bonus question, involving calculating the value of T_0 that minimizes pseudo-regret over all values of Δ, will receive full credit of 5 bonus points if any attempt was made at it, whether correct or not.

Problem 3 (The successive arm elimination algorithm) 20 points

(a) (4pts)

- 0pts: if no meaningful work is given.
- 1pts: if the right first step is considered, i.e. applying Hoeffding's inequality.
- 2pt: if partial steps are provided, e.g. the correct use of the Hoeffding's inequality for each value of t and each $i = 1, 2$, but the union bound is not correctly applied.
- 3pts: if the answer is mostly correct, but have some minor errors (e.g. incorrect constants).
- 4pts: for fully correct answer.

(b) (4pts) [The correct answer reasons that the eliminated arm must be the suboptimal arm 1.]

- 0pts: if no meaningful work is given.
- 1pts: if the right first step is considered, i.e. mentioning in your solution (without justification) that the eliminated arm is arm 1 under the good event.
- 2pts: if the correct relationship between the confidence region of the two arms is given, but this is not used to justify that $\mu_j < \mu_{j'}$ (where j is the eliminated arm).
- 3pts: if the justification that $\mu_j < \mu_{j'}$ (where j is the eliminated arm) is essentially correct with only minor errors (e.g. $\mu_j \leq \mu_{j'}$ instead of $\mu_j < \mu_{j'}$).
- 4pts: for the fully correct answer. *Note: as mentioned in the provided solutions, indexing errors in the definitions of $\mathsf{UCB}, \mathsf{LCB}$ will not be penalized.*

(c) (4pts) [The correct answer first derives an upper bound on $\mathsf{UCB}(2, \hat{t}-1) - \mathsf{LCB}(1, \hat{t}-1)$, then relates Δ to this.]

- 0pts: if no meaningful work is given.
- 1pts: if the right first step is considered, i.e. an attempt (whether incorrect or correct) to upper bound $\mathsf{UCB}(2, \hat{t}-1) - \mathsf{LCB}(1, \hat{t}-1)$.
- 2pts: if the correct relationship between $\mathsf{UCB}(2, \hat{t}-1) - \mathsf{LCB}(1, \hat{t}-1)$ is derived with minor errors, e.g. incorrect constants. Correct justification needs to be provided, i.e. that neither arm is eliminated.
- 2.5pts: if the correct relationship between $\mathsf{UCB}(2, \hat{t}-1) - \mathsf{LCB}(1, \hat{t}-1)$ is derived, but no further work is shown.
- 3pts: if the correct link between Δ and $\mathsf{UCB}(2, \hat{t}-1) - \mathsf{LCB}(1, \hat{t}-1)$ is established, but there are minor errors (e.g. incorrect constants, or usage of $N_{\hat{t}-1}(1) = \hat{t}$ instead of $\hat{t}-2$).
- 4pts: for fully correct answer. *Note: as mentioned in the provided solutions, indexing errors in the definitions of $\mathsf{UCB}, \mathsf{LCB}, N_{\hat{t}-1}$ will not be penalized as long as the correct answer is eventually derived.*

(d) (4pts) [The correct answer shows that the pseudo-regret is at most $\hat{t}/2$, and uses part (c) to derive an upper bound on $\hat{t}$.]

- 0pts: if no meaningful work is given.

- 1pts: if the right first step is considered, i.e. showing that the pseudo-regret is equal to $\hat{t}/2$, but no upper bound on $\hat{t}$ is provided.
- 2pts: if the upper bound of $\hat{t}$ is provided, but incorrectly connected to the definition of pseudo-regret.
- 3pts: if the proof is mostly correct, but contains some minor errors (e.g. pseudo-regret upper bounded by $(\hat{t}-2)/2$ instead of $\hat{t}/2$, or the factor $1/2$ is omitted).
- 4pts: for fully correct answer.

(e) (4pts) [The correct answer uses the law of total probability to evaluate the total pseudo-regret, and applies parts (d) and (a).]

- 0pts: if no meaningful work is given.
- 1pts: if the right first step is considered, i.e. expressing the pseudo-regret in terms of the law of total probability.
- 2pts: if one of the terms (related to the good or bad event) is correctly characterized
- 3pts: if both terms are essentially correct with only minor errors (e.g. $\frac{2}{T}$ is used for bad event probability instead of $\frac{4}{T}$, or $\frac{\hat{t}-2}{2}$ is used for good event contribution to pseudoregret instead of $\frac{\hat{t}}{2}$).
- 4pts: for the fully correct answer, i.e. both terms are correctly characterized. *In general, larger upper bounds than the ones provided still constitute valid solutions as long as the correct dependence on T and Δ (i.e. $\frac{\log T}{\Delta}$) is obtained and correct justification is provided.*