Advanced topics in ML and AGT **Fall Semester, 2018/19**

Homework 3: Dec 24,2018

Lecturer: Yishay Mansour

Homework number 3.

Generalized Minimax Theorem:
Let $X \subset R^d$ and $Y \subset R^d$ be convex compact sets. Let $f : X \times Y \to R$ be some differentiable function with bounded gradients, where $f(\cdot, y)$ is convex in its first argument for all fixed $y \in Y$, and $f(x, \cdot)$ is concave in its second argument for all fixed $x \in X$. Prove that

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) = \sup_{y \in Y} \inf_{x \in X} f(x, y).$$

Furthermore, give an efficient algorithm for finding an ϵ-optimal pair (x^*, y^*).

ϵ-greedy policy for stochastic MAB
An ϵ-greedy policy for MAB selects with probability ϵ a random action, and with probability $1 - \epsilon$ the action with the highest observed average reward. Consider the case of stochastic MAB, and derive a regret bound for the ϵ greedy policy as function of k, number of actions, T, number of time steps, and the parameter ϵ. Optimize the bound over the parameter ϵ to minimize the regret bound.

Lower Bounds stochastic MAB:
Let k be the number of actions. An instance of a stochastic MAB is $I = (p_1, \ldots, p_k)$, where $p_i \in (0, 1)$. The reward of action i under instance I is $Br(p_i)$, i.e., a Bernoulli random variable with probability p_i.

Consider an algorithm such that $E[Regret(T)] = O(C_{I,\alpha} T^{\alpha})$ for each problem instance I and each $\alpha > 0$. The "constant" $C_{I,\alpha}$ can depend on the problem instance I and the $\alpha > 0$, but not on time T.

Show that that for an arbitrary problem instance I there exists a time T_0 such that for any $T \geq T_0$, we have $E[Regret(T)] \geq C_I \log(T)$, for some constant C_I that depends on the problem instance, but not on time T.

Remark: The assumption is necessary to rule out trivial counterexamples: e.g., an algorithm which always play arm 1 will have zero regret on a problem instance for which arm 1 is the best arm. Thus, we cannot hope to prove the lower bound (or any non-trivial lower bound on regret that applies to every problem instance) without ruling out such trivial algorithms.

Swap regret and external regret
Show an example where the ratio between the external regret and swap regret can be unbounded (as function of the number of time steps T). (There is an example that uses only 3 actions, has zero external regret and swap regret linear in T.)

Clarification: You need to choose only a loss sequence and actions (no adversary or algorithm). Given the loss sequence and the selected actions, the external regret is the difference between the loss of the action sequence and that of the loss of the best single action. The swap regret is the difference between the loss of the action sequence and the loss of the best function F of swapping actions.

The homework is due in two weeks