MLT Homework 8

Ana Borovac
Bas Haver

November 11, 2018

Question 1

Let $\psi(\lambda) = \frac{\lambda^2}{2}$. The Legendre-Fenchel transform of ψ is given by

$$\psi^*(\epsilon) = \sup_{\lambda \in \mathbb{R}} \lambda\epsilon - \psi(\lambda).$$

Subquestion 1.1

$\psi^*(\epsilon) = \frac{\epsilon^2}{2}$.

Solution

Let define $\psi_1(\lambda)$:

$$\psi_1(\lambda) = \lambda\epsilon - \frac{\lambda^2}{2}$$

Furthermore:

$$\psi_1'(\lambda) = \epsilon - \lambda$$

So, the maximum is reached at:

$$\lambda = \epsilon \quad \Rightarrow \quad \psi_1(\epsilon) = \epsilon \cdot \epsilon - \frac{\epsilon^2}{2} = \frac{\epsilon^2}{2}$$

We can conclude:

$$\psi^*(\epsilon) = \psi_1(\epsilon) = \frac{\epsilon^2}{2}$$

Subquestion 1.2

$(\psi^*)^{-1}(z) = \pm\sqrt{2z}$.

Solution

From previous point we know:

$$\psi^*(\epsilon) = \frac{\epsilon^2}{2}$$

It follows:

$$\begin{aligned} z &= \frac{\epsilon^2}{2} \\ 2z &= \epsilon^2 \\ \epsilon &= \pm\sqrt{2z} \end{aligned}$$

So:

$$(\psi^*)^{-1}(z) = \pm\sqrt{2z}$$

Question 2

The Blooper Reel

Subquestion 2.1

Deterministic fails for Adversarial Bandits *Show that any deterministic algorithm (UCB included) has linear regret in the adversarial bandit setting. Hint: you can use the argument on the top of page 23.*

Question 3

We consider an adversarial bandit model with K^2 arms indexed by $i \in [K]$ and $j \in [K]$. For each arm (i, j), the loss at time t is $a_t^i + b_t^j$, where $a_t^i \in [0, 1]$ and $b_t^j \in [0, 1]$ are chosen by the adversary before the start of the interaction. Then each round the learner picks an arm $(I_t, J_t) \in [K]^2$ and observes $a_t^{I_t}$ and $b_t^{J_t}$ separately (and incurs their sum as the loss).

Subquestion 3.1

Consider running a single instance of EXP3 on all K^2 arms (with loss range $[0, 2]$). Show that the expected pseudo-regret compared to the best arm (i^*, j^*) is bounded by

$$\bar{R}_n \leq 2\sqrt{2nK^2 \ln(K^2)}$$

Solution

$$\bar{R}_n = \mathbb{E}\sum_{t=1}^{T} l_{t,I_t,J_t} - \min_{k_1,k_2}\sum_{t=1}^{T} l_{t,k_1,k_2}$$

$$\widetilde{l}_{t,i,j} = \frac{l_{t,i,j}}{p_{t,i,j}}\mathbb{1}_{I_t=i,J_t=j} \quad \Rightarrow \quad \mathbb{E}\left[\widetilde{l}_{t,i,j}\right] = l_{t,i,j}$$

EXP3 algorithm:
learning rate η
set $p_{t,i,j} = \dfrac{exp(\eta \sum_{s=1}^{t-1} \widetilde{l}_{t,i,j})}{norm}$
sample $I_t \sim p_t$, $J_t \sim p_t$
set $\widetilde{l}_{t,i,j} = \frac{l_{t,i,j}}{p_{t,i,j}}\mathbb{1}_{I_t=i,J_t=j}$
Theorem: $\bar{R}_T \leq \sqrt{2TK^2 \ln K^2}$
Lemma: $\bar{R}_T \leq \frac{K^2 T\eta}{2} + \frac{\log K^2}{\eta}$
Hadge analysis on $\widetilde{l}$

Subquestion 3.2

Now we will use the a_t^i and b_t^j observations separately. Consider running two K-arm instances of EXP3, one with $i \to a_t^i$ as the loss and one with $j \to b_t^j$ as the loss. Have the first algorithm control I_t and the second J_t. Show that the overall expected pseudo-regret is bounded by

$$\bar{R}_n \leq 2\sqrt{2nK\ln K}.$$

Solution

From the lectures we know that the regret of one K-arm EKP3 algorithm is bounded with $\sqrt{2nK\ln K}$. So, in our case:

$$\bar{R}_n = \bar{R}_n^1 + \bar{R}_n^2 \leq \sqrt{2nK\ln K} + \sqrt{2nK\ln K} = 2\sqrt{2nK\ln K}$$