Mehryar Mohri
Advanced Machine Learning 2018
Courant Institute of Mathematical Sciences
Homework assignment 1
February 20, 2018
Due: March 06, 2018

**A. Exponentially Weighted algorithm**

We consider the Exponentially Weighted algorithm and adopt the notation and setup discussed in class. Let $N_L$ be the number of experts with cumulative loss at most $L > 0$ at time $T$: $N_L = |\{i \in [N] \colon \sum_{t=1}^{T} L(\widehat{y}_{t,i}, y_t) \leq L\}|$.

1. Show that the following inequality holds for the cumulative loss of the algorithm, for any $\eta > 0$:

$$\sum_{t=1}^{T} L(\widehat{y}_t, y_t) \leq L + \frac{1}{\eta} \log \frac{N}{N_L} + \frac{\eta T}{8}.$$

2. Suppose $L$ is very close to $\min_{i=1}^{N} L_{T,i}$ and that $N_L$ is very large, say $N_L = N/2$. What does the bound show?

**B. Games**

1. Find all pure and mixed Nash equilibria of the following game.

| | L | R |
|---|---|---|
| U | (0, 0) | (6, 2) |
| D | (2, 6) | (5, 5) |

2. Can you find a correlated equilibrium for which the sum of the players' payoffs is more favorable than that of any Nash equilibrium?

3. Same two questions for the following game.

| | L | C | R |
|---|---|---|---|
| U | (1, 1) | (2, 4) | (4, 2) |
| M | (4, 2) | (1, 1) | (2, 4) |
| D | (2, 4) | (4, 2) | (1, 1) |

**C. Alternative proof of the theorem of Nash**

In class we gave a full proof of the theorem of Nash. Here, we will give an alternative proof using Bregman divergences. We will adopt the notation and terminology introduced in class.

For any $k \in [p]$, let $F_k$ be a strictly convex nd differentiable function defined over an open and convex set $C_k$ containing the simplex $\Delta(\mathcal{A}_k)$. We will denote by $\mathsf{B}_k$ the Bregman divergence associated to $F_k$.

1. For any $\mathsf{p} \in \Omega = \times_{k=1}^{p} \Delta(\mathcal{A}_k)$ and $k \in [p]$, we define the function $\mathsf{q} \mapsto \Psi_k(\mathsf{q}, \mathsf{p})$ over $\Delta(\mathcal{A}_k)$ by

$$\Psi_k(\mathsf{q}, \mathsf{p}) = -u_k(\mathsf{q}, \mathsf{p}_{-k}) + \mathsf{B}_k(\mathsf{q} \,\|\, \mathsf{p}_k).$$

   Prove that, for any $\mathsf{p} \in \Omega$, $\min_{q \in \Delta(\mathcal{A}_k)} \Psi_k(\mathsf{q}, \mathsf{p})$ is attained.

2. Prove that the minimizer of $\Psi_k(\mathsf{q}, \mathsf{p})$ over $\Delta(\mathcal{A}_k)$ is unique (*hint*: show that $\mathsf{q} \mapsto \Psi_k(\mathsf{q}, \mathsf{p})$ is strictly convex).

3. Let $f \colon \Omega \to \Omega$ be the function defined for any $\mathsf{p} \in \Omega$ by $f(\mathsf{p}) = (\mathsf{q}_1, \ldots, \mathsf{q}_p)$, with $\mathsf{q}_k = \operatorname{argmin}_{\mathsf{q} \in \Delta(\mathcal{A}_k)} \Psi_k(\mathsf{q}, \mathsf{p})$. Assume that, for any $\mathsf{p} \in \Omega$, $\operatorname{argmin}_{\mathsf{q} \in \Delta(\mathcal{A}_k)} \Psi_k(\mathsf{q}, \mathsf{p})$ is a continuous function of $\mathsf{p}$.

   Show that $f$ is well defined and show that $f$ admits a fixed-point.

4. Let $\mathsf{p} \in \Omega$ be a fixed-point of $f$. Show that for any $k \in [p]$ and $q \in \Delta(\mathcal{A}_k)$, the following inequality holds: $\Psi_k(\mathsf{q}, \mathsf{p}) \geq \Psi_k(\mathsf{p}_k, \mathsf{p})$ (*hint*: prove that the right-derivative of the function $J$ defined over $[0, 1]$ by $J(\alpha) = \Psi_k(\alpha \mathsf{q} + (1 - \alpha)\mathsf{p}_k, \mathsf{p})$ is non-negative). Prove the theorem of Nash.

5. Show that the function $F_k$ defining the Bregman divergence $\mathsf{B}_k(\mathsf{q} \,\|\, \mathsf{p}_k) = \frac{1}{2}\|\mathsf{q} - \mathsf{p}_k\|_2^2$ satisfies the assumptions.

6. Let $n_k$ be the cardinality of $\mathcal{A}_k = \{a_1, \ldots, a_{n_k}\}$ and let $\mathsf{v} \in \mathbb{R}^{n_k}$ be the vector whose $j$th coordinate is $u_k(a_j, \mathsf{p}_{-k})$, $j \in [n_k]$. Prove that

$$\operatorname*{argmin}_{\mathsf{q} \in \Delta(\mathcal{A}_k)} \Psi_k(\mathsf{q}, \mathsf{p}) = \operatorname*{argmin}_{\mathsf{q} \in \mathcal{A}_k} \frac{1}{2}\|\mathsf{q} - (\mathsf{p}_k + \mathsf{v})\|_2^2.$$

7. [bonus question] Prove that $\mathsf{p} \mapsto \operatorname{argmin}_{\mathsf{q} \in \Delta(\mathcal{A}_k)} \Psi_k(\mathsf{q}, \mathsf{p})$ is continuous.