CS7545, Spring 2023: Machine Learning Theory - Homework #4

Jacob Abernethy, Zihao Hu, and Guanghui Wang Due: Tuesday, April 25 at 11:59 p.m.

Homework Policy: *The due is at Tuesday 4/25, but everyone gets a free extension to 5/2.* Working in groups is fine, but *every student* must submit their own writeup. Please write the members of your group on your solutions. There is no strict limit to the size of the group but we may find it a bit suspicious if there are more than 4 to a team. Questions labelled with **(Challenge)** are not strictly required, but you'll get some participation credit if you have something interesting to add, even if it's only a partial answer.

1) **Generalized Minimax Theorem.** Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be convex compact sets. Let $f : X \times Y \to \mathbb{R}$ be some differentiable function with bounded gradients, where $f(\cdot, \mathbf{y})$ is convex in its first argument for all fixed $\mathbf{y}$, and $f(\mathbf{x}, \cdot)$ is concave in its second argument for all fixed $\mathbf{x}$. An ϵ-optimal point $(\mathbf{x}^*, \mathbf{y}^*)$ satisfies

$$\sup_{\mathbf{y}\in Y} f(\mathbf{x}^*, \mathbf{y}) - \inf_{\mathbf{x}\in X} f(\mathbf{x}, \mathbf{y}^*) \leq \epsilon.$$

- Prove that

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

- Give an efficient algorithm which finds a $O(T^{-1/2})$-optimal $(\mathbf{x}^*, \mathbf{y}^*)$ in T iterations.
- Assume $f(\cdot, \cdot)$ is strongly convex in its first argument and strongly concave in its second argument. Find a $O(\frac{\log T}{T})$-optimal solution after T iterations.

2) **Minimax.** As in Von Neumann's minimax theorem, we can easily verify that the weak duality holds.

Similarly to the proof for strong duality we saw in class, suppose the players P_1 and P_2 play a game, learning from each other using OGD. This is possible because our function is differentiable. The regret is sub-linear in T due to our assumption that G and D are bounded. The player P_1 receives a sequence of functions $f(\cdot, y^0), \ldots, f(\cdot, y^{t-1})$ and chooses x^t, and the player P_2 receives a sequence of functions $-f(x^0, \cdot), \ldots, -f(x^{t-1}, \cdot)$ and chooses y^t.

Using the regret bound of OGD,

$$\frac{1}{T}\sum_{t=1}^{T} f(x^t, y^t) \leq \frac{1}{T}\min_x \sum_{t=1}^{T} f(x, y^t) + \frac{1}{T} O(GD\sqrt{T}) \leq \min_x f(x, \hat{y}^T) + \epsilon(T)$$

where $\hat{y}^T = \frac{1}{T}\sum_{t=1}^{T} y^t$ and $\epsilon(T)$ is some positive sub-linear function of T. The inequality holds because f is concave in the second argument. Similarly,

$$\frac{1}{T}\sum_{t=1}^{T} -f(x^t, y^t) \leq \frac{1}{T}\min_y - \sum_{t=1}^{T} f(x^t, y) + \frac{1}{T} O(GD\sqrt{T}) \leq \min_y -f(\hat{x}^T, y) + \epsilon(T)$$

which implies

$$\frac{1}{T}\sum_{t=1}^{T} f(x^t, y^t) \geq \max_y f(\hat{x}^T, y) - \epsilon(T)$$

where $\hat{x}^T = \frac{1}{T}\sum_{t=1}^T x^t$. The inequality holds because $-f$ is concave in the first argument.

Define $V(T) := \frac{1}{T}\sum_{t=1}^T f(x^t, y^t)$ and now we connect the two inequalities:

$$\min_x \max_y f(x,y) - \epsilon(T) \leq \max_y f(\hat{x}^T, y) - \epsilon(T) \leq V(T) \leq \min_x f(x, \hat{y}^T) + \epsilon(T) \leq \max_y \min_x f(x,y) + \epsilon(T)$$

So,

$$\min_x \max_y f(x,y) \leq \max_y \min_x f(x,y) + 2\epsilon(T)$$

.

To find a ϵ-optimal pair, run the OGD as described above for a large enough number of iterations, and the average move across time $\hat{x}^T, \hat{y}^T$ is an ϵ-optimal pair. For the strongly-convex strongly-concave case. We only need to use OGD for strongly convex functions and plug in the $O(\log T)$ regret.