Mehryar Mohri
Advanced Machine Learning 2016
Courant Institute of Mathematical Sciences
Homework assignment 1
March 7th, 2015
Due: March 21, 2015

A. Deep ensembles

In class, we presented a guarantee for the SRM method. This problem suggests you to derive a similar guarantee for deep ensembles or the principle of voted risk minimization (VRM). We will largely adopt the notation used in class. Let $\mathcal{X}$ denote the input space and let $H_1, \ldots, H_p$ families of functions mapping from $\mathcal{X}$ to $\mathbb{R}$. $\mathcal{F}$ is the convex hull of the union of these families:

$$\mathcal{F} = \operatorname{conv}\left(\bigcup_{k=1}^{p} H_k\right) = \left\{\sum_{j=1}^{N} \alpha_j h_j \colon \alpha_j \geq 0, \sum_{j=1}^{N} \alpha_j \leq 1\right\}.$$

Let F be the objective function defined for all $f = \sum_{j=1}^{N} \alpha_j h_j \in \mathcal{F}$ by

$$F(f) = \widehat{R}_{S,\rho}(f) + \frac{4}{\rho} \sum_{j=1}^{N} \alpha_j \mathfrak{R}_m(H_{k(j)}),$$

where $H_{k(j)}$ is the least complex family containing h_j. Define the VRM solution as the function f_{VRM} minimizing F:

$$f_{\text{VRM}} = \operatorname*{argmin}_{f \in \mathcal{F}} F(f).$$

Use Corollary 1 and Corollary 2 (see appendix) to show the following result. For any $\delta > 0$, with probability at least $1 - \delta$ over the choice of a sample S of size m drawn i.i.d. according to $\mathcal{D}^m$, the following inequality holds for f_{VRM}:

$$R(f_{\text{VRM}}) \leq \inf_{f \in \mathcal{F}} \left(R_\rho(f) + \frac{8}{\rho} \sum_{j=1}^{N} \alpha_j \mathfrak{R}_m(H_{k(j)}) + \frac{4}{\rho} \sqrt{\frac{\log p}{m}} \left[1 + \sqrt{\left\lceil \log\left[\frac{\rho^2 m}{\log p}\right] \right\rceil} \right] \right) + \sqrt{\frac{2 \log \frac{4}{\delta}}{m}},$$

with $R_\rho(f) = \mathrm{E}[\Phi_\rho(yf(x))]$, where Φ_ρ is the margin loss function: $\Phi_\rho(x) = \min\left(1, \max\left(0, 1 - \frac{x}{\rho}\right)\right)$ for all $x \in \mathbb{R}$.

Note that the bound applies in particular with the right-hand side function chosen to be f^*, the one minimizing F: $f_{\mathrm{VRM}} = \operatorname{argmin}_{f \in \mathcal{F}} F(f)$.

B. Zero-sum games

In class, we gave a proof of von Neumann's minimax theorem by assuming that one of the players was using the RWM algorithm. Consider the scenario where both players use RWM at each round, which we alluded to in class. Prove von Neumann's minimax theorem using that scenario, proceeding as follows.

1. Assume without loss of generality that $u_1 \leq 1$. Let p_t be the distribution defined at the tth round by the row player and q_t the one for the column player. Show that

$$\max_{\mathsf{p}} \frac{1}{T} \sum_{t=1}^{T} \mathop{\mathrm{E}}_{\substack{a_1 \sim \mathsf{p} \\ a_2 \sim \mathsf{q}_t}} [u_1(\mathbf{a})] - \frac{R_T}{T} \leq \min_{\mathsf{q}} \frac{1}{T} \sum_{t=1}^{T} \mathop{\mathrm{E}}_{\substack{a_1 \sim \mathsf{p}_t \\ a_2 \sim \mathsf{q}}} [u_1(\mathbf{a})] + \frac{R_T}{T}.$$

2. Use the previous inequality and the regret bound for RWM to prove von Neumann's minimax theorem.

C. Correlated Equilibria

Consider the game defined by the following matrix:

	A	B
A	(8, 8)	(1, 9)
B	(9, 1)	(0, 0)

1. Which are the pure Nash equilibria for this game?

2. Find a mixed Nash equilibrium. Which is the expected payoff for the row player?

3. Suppose now that a correlation device draws each of (A, A), (A, B), (B, A) with equal probability. Prove that this defines a correlated equilibrium. What is the expected payoff? How does it compare to the expected payoff for a mixed Nash equilibrium found in the previous question?

A Appendix

Corollary 1 *Assume $p > 1$. Fix $\rho > 0$. Then, for any $\delta > 0$, with probability at least $1 - \delta$ over the choice of a sample S of size m drawn i.i.d. according to $\mathcal{D}^m$, the following inequality holds for all $f = \sum_{j=1}^{N} \alpha_j h_j \in \mathcal{F}$:*

$$R(f) \leq \widehat{R}_{S,\rho}(f) + \frac{4}{\rho} \sum_{j=1}^{N} \alpha_j \mathfrak{R}_m(H_{k(j)}) + \frac{2}{\rho} \sqrt{\frac{\log p}{m}} \left[1 + \sqrt{\left\lceil \log \left[\frac{\rho^2 m}{\log p}\right] \right\rceil}\right] + \sqrt{\frac{\log \frac{2}{\delta}}{2m}}.$$

Corollary 2 *Assume $p > 1$. Fix $\rho > 0$. Then, for any $\delta > 0$, with probability at least $1 - \delta$ over the choice of a sample S of size m drawn i.i.d. according to $\mathcal{D}^m$, the following inequality holds for all $f = \sum_{j=1}^{N} \alpha_j h_j \in \mathcal{F}$:*

$$\widehat{R}_{S,\rho}(f) \leq R_\rho(f) + \frac{4}{\rho} \sum_{j=1}^{N} \alpha_j \mathfrak{R}_m(H_{k(j)}) + \frac{2}{\rho} \sqrt{\frac{\log p}{m}} \left[1 + \sqrt{\left\lceil \log \left[\frac{\rho^2 m}{\log p}\right] \right\rceil}\right] + \sqrt{\frac{\log \frac{2}{\delta}}{2m}}.$$