Supplement: A Robust Zero-Sum Game Framework for Pool-based Active Learning

A Proof of Theorem 1

The first part of the Theorem follows Theorem 1 in [8], and the second part follows Corollary 3.2 in [8].

B Proof of Theorem 3 and Step Size Setting

Proof. Following Lemma 1 of [26], for any $\mathbf{p} \in \Delta_n$ and $\mathbf{w} \in \Omega$ we have

$$\frac{1}{T}\mathrm{E}\left[\sum_{t=1}^{T} \mathbf{p}^\top \bar{\ell}(\mathbf{w}_t) - \mathbf{p}_t^\top \bar{\ell}(\mathbf{w})\right] \leq \frac{\|\mathbf{w} - \mathbf{w}_1\|_2^2}{T\eta} + \frac{\eta}{2}G^2 + \frac{1}{T}\mathrm{E}\left[\sum_{t=1}^{T} \bar{\ell}(\mathbf{w}_t)^\top (\mathbf{p} - \mathbf{p}_t)\right]$$

Next, we bound the last term in the above inequality.

$$\mathrm{E}\left[\sum_{t=1}^{T} \bar{\ell}(\mathbf{w}_t)^\top (\mathbf{p} - \mathbf{p}_t)\right] = \mathrm{E}\left[\sum_{t=1}^{T} \hat{\mathbf{v}}_t^\top (\mathbf{p} - \mathbf{p}_t)\right] + \mathrm{E}\left[\sum_{t=1}^{T} (\bar{\ell}(\mathbf{w}_t) - \hat{\mathbf{v}}_t)^\top (\mathbf{p} - \mathbf{p}_t)\right]$$

Following the analysis of mirror descent on $\mathbf{p}_t$ (e.g., Lemma 2 in [26]), we have

$$\mathrm{E}\left[\sum_{t=1}^{T} \hat{\mathbf{v}}_t^\top (\mathbf{p} - \mathbf{p}_t)\right] \leq \frac{D(\mathbf{p}, \mathbf{p}_1)}{\alpha} + \alpha \sum_{t=1}^{T} \|\bar{\mathbf{v}}_t\|_\infty^2 \leq \frac{D(\mathbf{p}, \mathbf{p}_1)}{\alpha} + T\alpha M^2$$

In addition,

$$\begin{aligned}\mathrm{E}\left[\sum_{t=1}^{T} (\bar{\ell}(\mathbf{w}_t) - \hat{\mathbf{v}}_t)^\top (\mathbf{p} - \mathbf{p}_t)\right] &\leq \mathrm{E}\left[\sum_{t=1}^{T} MR\|\mathbf{w}_t - \mathbf{w}_*\|_2\right] \\ &\leq \mathrm{E}\left[\sum_{t=1}^{T} MR(\|\mathbf{w}_t - \widehat{\mathbf{w}}\|_2 + \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2)\right]\end{aligned}$$

$$
\begin{aligned}
&\leq TMR \cdot O(1/n^{\beta}) + \mathrm{E}\left[\sum_{t=1}^{T} MR\hat{c}(\mathcal{L}(\mathbf{w}_t) - \mathcal{L}(\widehat{\mathbf{w}}))\right] \\
&\leq TMR \cdot O(1/n^{\beta}) + TMR\hat{c}\mathrm{E}[(\mathcal{L}(\mathbf{w}_\tau) - \mathcal{L}(\widehat{\mathbf{w}}))]
\end{aligned}
$$

where the first inequality uses that $1/\exp(-Y\mathbf{w}^\top\mathbf{x})$ is $R/4$-Lipchitz continuous function, $\bar{\ell}(\mathbf{w}_t;\mathbf{x}_i) \in [0, M]$ and $\|\mathbf{p} - \mathbf{p}_t\|_1 \leq 2$. Combining the above inequalities together, we have

$$
\begin{aligned}
\mathrm{E}[\mathbf{p}^\top\bar{\ell}(\mathbf{w}_\tau)] - \max_{\mathbf{p}\in\Delta_n} \mathbf{p}^\top\bar{\ell}(\widehat{\mathbf{w}}) \leq \frac{4r^2}{T\eta} + \frac{\eta}{2}G^2 + \frac{D(\mathbf{p},\mathbf{p}_1)}{\alpha T} \\
+ \alpha M^2 + MR \cdot O(1/n^{\beta}) + MR\hat{c}\mathrm{E}[(\mathcal{L}(\mathbf{w}_\tau) - \mathcal{L}(\widehat{\mathbf{w}}))]
\end{aligned}
$$

Let $\mathbf{p}$ be the vector that maximizes $\mathbf{p}^\top\bar{\ell}(\mathbf{w}_\tau)$, and suppose $D(\mathbf{p},\mathbf{p}_1) \leq \mathcal{D}$, $\eta = 2\sqrt{2}r/(G\sqrt{T})$ and $\alpha = \sqrt{\mathcal{D}/(M^2T)}$, then

$$
\mathrm{E}[(\mathcal{L}(\mathbf{w}_\tau) - \mathcal{L}(\widehat{\mathbf{w}}))] \leq \frac{4\sqrt{2}rG}{\sqrt{T}} + 4\sqrt{\mathcal{D}}\frac{M}{\sqrt{T}} + MR \cdot O\left(\frac{1}{n^{\beta}}\right)
$$

For constrained problems, we have $\mathcal{D} = \rho/n$. For regularized problems, we can use the bound $D(\mathbf{p},\mathbf{p}_1) \leq \log(n)$ for KL divergence and $D(\mathbf{p},\mathbf{p}_1) \leq 2$ for Euclidean divergence. □

C Dataset Statistics

Table 1: Statistics of Datasets

Dataset	# Feature	# Training	# Testing
svmguide3	22	1243	41
breast-cancer	9	200	77
twonorm	20	400	7000
ringnorm	20	400	7000
flare solar	9	666	400
heart	13	170	100
german	20	700	300
diabetis	8	468	300
duke breast-cancer	7129	38	4
madelon	500	2000	600
MNIST	$28 * 28$	55000	10000
CIFAR-10	$32 * 32 * 3$	50000	10000