Supplemental Material

A Proofs

When bounding the Rademacher complexity for Lipschitz continuous loss classes (such as the hinge loss or the squared loss), the following lemma is often very helpful.

Lemma A.1 (Talagrand's lemma [41])**.** *Let $l : \mathbb{R} \to \mathbb{R}$ be a loss function that is L-Lipschitz continuous and $l(0) = 0$. Let $\mathcal{F}$ be a hypothesis class of real-valued functions and denote its loss class by $\mathcal{G} := l \circ \mathcal{F}$. Then the following inequality holds:*

$$R_n(\mathcal{G}) \leq 2LR_n(\mathcal{F}).$$

We can use the above result to prove Lemma 3.

Proof of Lemma 3. Since the LATENTSVDD loss function is 1-Lipschitz with $l(0) = 0$, by Lemma A.1, it is sufficient to bound $R(\mathcal{F}_{\text{SVDD}}(\boldsymbol{z}))$. To this end, it holds

$$\begin{aligned} R(\mathcal{F}_{\text{SVDD}}(\boldsymbol{z})) &\overset{\text{def.}}{=} \mathbb{E}\Big[\sup_{\boldsymbol{c},\Omega:0\leq\|\boldsymbol{c}\|^2+\Omega\leq\lambda} \frac{1}{n}\sum_{i=1}^n \sigma_i\big(\Omega \\ &\quad + 2\langle \boldsymbol{c}, \Psi(\boldsymbol{x}_i,\boldsymbol{z})\rangle - \|\Psi(\boldsymbol{x}_i,\boldsymbol{z})\|^2\big)\Big] \\ &\leq \mathbb{E}\left[\sup_{\Omega:-\lambda\leq\Omega\leq\lambda} \frac{1}{n}\sum_{i=1}^n \sigma_i\Omega\right] \\ &\quad + 2\mathbb{E}\left[\sup_{\boldsymbol{c}:\|\boldsymbol{c}\|^2\leq\lambda} \frac{1}{n}\sum_{i=1}^n \sigma_i\left(\langle \boldsymbol{c}, \Psi(\boldsymbol{x}_i,\boldsymbol{z})\rangle\right)\right] \\ &\quad + \underbrace{\mathbb{E}\left[-\frac{1}{n}\sum_{i=1}^n \sigma_i\|\Psi(\boldsymbol{x}_i,\boldsymbol{z})\|^2\right]}_{=0 \text{ (by symmetry of } \sigma_i)}. \end{aligned} \quad \text{(A.3.1)}$$

Note that the term to the right is zero because the Rademacher variables are random signs, independent of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. The term to the left can be bounded as follows:

$$\begin{aligned} \mathbb{E}\left[\sup_{\Omega:-\lambda\leq\Omega\leq\lambda} \frac{1}{n}\sum_{i=1}^n \sigma_i\Omega\right] &= \lambda\mathbb{E}\left[\left| \frac{1}{n}\sum_{i=1}^n \sigma_i\right| \right] \\ &\overset{(*)}{\leq} \lambda\sqrt{\mathbb{E}\left[\frac{1}{n^2}\sum_{i,j=1}^n \sigma_i\sigma_j\right]} \\ &= \frac{\lambda}{\sqrt{n}}. \end{aligned} \quad \text{(A.3.2)}$$

where for $(*)$ we employ Jensen's inequality. Moreover, applying the Cauchy-Schwarz inequality and Jensen's inequality, respectively, we obtain

$$\begin{aligned} &\mathbb{E}\left[\sup_{\boldsymbol{c}:\|\boldsymbol{c}\|^2\leq\lambda} \frac{1}{n}\sum_{i=1}^n \sigma_i\left(\langle \boldsymbol{c}, \Psi(\boldsymbol{x}_i,\boldsymbol{z})\rangle\right)\right] \\ &\overset{\text{C.-S.}}{\leq} \mathbb{E}\left[\sup_{\boldsymbol{c}:\|\boldsymbol{c}\|^2\leq\lambda} \|\boldsymbol{c}\| \left\| \frac{1}{n}\sum_{i=1}^n \sigma_i\Psi(\boldsymbol{x}_i,\boldsymbol{z})\right\| \right] \\ &\overset{\text{Jensen}}{\leq} \sqrt{\lambda\mathbb{E}\left[\frac{1}{n^2}\sum_{i,j=1}^n \sigma_i\sigma_j\langle\Psi(\boldsymbol{x}_i,\boldsymbol{z}),\Psi(\boldsymbol{x}_j,\boldsymbol{z})\rangle\right]} \\ &= \sqrt{\lambda\frac{1}{n^2}\sum_{i=1}^n \|\Psi(\boldsymbol{x}_i,\boldsymbol{z})\|^2} \\ &\leq B\sqrt{\frac{\lambda}{n}} \end{aligned} \quad \text{(A.3.3)}$$

because $\mathbb{P}(\|\Psi(\boldsymbol{x}_i,\boldsymbol{z})\| \leq B) = 1$. Hence, inserting the results (A.3.2) and (A.3.3) into (A.3.1), yields the claimed result, that is,

$$\begin{aligned} R(\mathcal{G}_{\text{SVDD}}(\boldsymbol{z})) &\overset{\text{Lemma A.1}}{\leq} R(\mathcal{F}_{\text{SVDD}}(\boldsymbol{z})) \\ &\leq \frac{\lambda}{\sqrt{n}} + B\sqrt{\frac{\lambda}{n}} = \frac{\lambda + B\sqrt{\lambda}}{\sqrt{n}}. \end{aligned} \quad \text{(A.3.4)}$$

□

Next, we invoke the following result, taken from [23] (Lemma 8.1).

Lemma A.2. *Let $\mathcal{F}_1, \ldots, \mathcal{F}_l$ be hypothesis sets in $\mathbb{R}^{\mathcal{X}}$, and let $\mathcal{F} := \{\max(f_1, \ldots, f_l) : f_i \in \mathcal{F}_i, i \in \{1, \ldots, l\}\}$. Then,*

$$R_n(\mathcal{F}) \leq \sum_{j=1}^l R_n(\mathcal{F}_j).$$

Sketch of proof [23]. The idea of the proof is to write $\max(h_1, h_2) = \frac{1}{2}(h_1 + h_2 + |h_1 - h_2|)$, and then to show that

$$\mathbb{E}\left[\sup_{h_1\in\mathcal{F}_1,h_2\in\mathcal{F}_2} \frac{1}{n}\sum_{i=1}^n |h_1(x_i) - h_2(x_i)|\right] \leq R_n(\mathcal{F}_1)+R_n(\mathcal{F}_2).$$

This proof technique also generalizes to $l > 2$. □

We can use Lemma A.2 and Lemma 3, to conclude the main theorem of this paper, that is, Theorem 2, which establishes generalization guarantees of the usual order $O(1/\sqrt{n})$ for the proposed LATENTSVDD method.

Proof of Theorem 2. First observe that, because l is 1-Lipschitz,

$$R_n(\mathcal{G}_{\text{LATENTSVDD}}) \leq R_n(\mathcal{F}_{\text{LATENTSVDD}}).$$

Next, note that we can write

$$R_n(\mathcal{F}_{\text{LATENTSVDD}}) \;=\; \left\{\max_{\boldsymbol{z}\in\mathcal{Z}}(f_{\boldsymbol{z}}) : f_{\boldsymbol{z}} \in \mathcal{F}_{\text{SVDD}}(\boldsymbol{z})\right\}.$$

Thus, by Lemma 2 and Lemma 4,

$$\begin{aligned} R_n(\mathcal{F}_{\text{LATENTSVDD}}) \;&\leq\; |\mathcal{Z}| \max_{z\in\mathcal{Z}} R_n(\mathcal{F}_{\text{SVDD}}(\boldsymbol{z})) \\ &\leq\; |\mathcal{Z}| \frac{\lambda + B\sqrt{\lambda}}{\sqrt{n}}. \end{aligned}$$

Moreover, observe that the loss function in the definition of $\mathcal{G}_{\text{LATENTSVDD}}$ can only range in the interval $[0, B]$. Thus, Theorem 2 in the main paper gives the claimed result, that is,

$$\begin{aligned} \mathbb{E}[\widehat{g}_n] - \mathbb{E}[g^*] \;&\leq\; 4R_n(\mathcal{G}_{\text{LATENTSVDD}}) + B\sqrt{\frac{2\log(2/\delta)}{n}} \\ &\leq\; 4|\mathcal{Z}| \frac{\lambda + B\sqrt{\lambda}}{\sqrt{n}} + B\sqrt{\frac{2\log(2/\delta)}{n}}. \end{aligned}$$

$\square$