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 \( R(\mathcal{F}_{\text{SVDD}}(z)) \). To this end, it holds

\[
R(\mathcal{F}_{\text{SVDD}}(z)) \leq \sup_{c: \forall \Omega \in \mathcal{F}_{\text{SVDD}}(z)} \frac{1}{n} \sum_{i=1}^{n} \sigma_i(\Omega) + 2\langle c, \Psi(x_i, z) \rangle = \left\| \sum_{i=1}^{n} \sigma_i(\Omega) \right\|
\]

which can be bounded as follows:

\[
\mathbb{E} \left[ \sup_{\Omega : -\lambda \leq \Omega \leq \lambda} \frac{1}{n} \sum_{i=1}^{n} \sigma_i(\Omega) \right] \leq \lambda \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} \sigma_i(\Omega) \right] \leq \frac{\lambda}{\sqrt{n}}.
\]

Note that the term to the right is zero because the Rademacher variables are random signs, independent of \( x_1, \ldots, x_n \). The term to the left can be bounded as follows:

\[
\mathbb{E} \left[ \sup_{\Omega : -\lambda \leq \Omega \leq \lambda} \frac{1}{n} \sum_{i=1}^{n} \sigma_i(\Omega) \right] \leq \lambda \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} \sigma_i(\Omega) \right] \leq \frac{\lambda}{\sqrt{n}}.
\]

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

\[
R(\mathcal{G}_{\text{SVDD}}(z)) \leq \frac{\lambda}{\sqrt{n}} + B \sqrt{\frac{\lambda}{n}} = \frac{\lambda + B\sqrt{\lambda}}{\sqrt{n}}.
\]

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 \( \mathbb{R}^N \), 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
\[
R_n(\mathcal{F}_{\text{LATENTSVDD}}) = \left\{ \max_{z \in \mathcal{Z}} (f_z) : f_z \in \mathcal{F}_{\text{SVDD}}(z) \right\}.
\]

Thus, by Lemma 2 and Lemma 4,
\[
R_n(\mathcal{F}_{\text{LATENTSVDD}}) \leq |\mathcal{Z}| \max_{z \in \mathcal{Z}} R_n(\mathcal{F}_{\text{SVDD}}(z))
\leq |\mathcal{Z}| \frac{\lambda + B \sqrt{\lambda}}{\sqrt{n}}.
\]

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,
\[
\mathbb{E}[g_n] - \mathbb{E}[g^*] \leq 4R_n(\mathcal{G}_{\text{LATENTSVDD}}) + B \sqrt{\frac{2 \log(2/\delta)}{n}}
\leq 4|\mathcal{Z}| \frac{\lambda + B \sqrt{\lambda}}{\sqrt{n}} + B \sqrt{\frac{2 \log(2/\delta)}{n}}.
\]