Appendix A. Discussion

We can extend the definition of robustness to triplet-based metric learning algorithms, i.e., we take the admissible triplet set $\text{trip}_{\mathbf{s}}$ of s such that $(s_1, s_2, s_3) \in \text{trip}_{\mathbf{s}}$. The robustness property means s_1 and s_2 share the same label while s_1 and s_3 have different ones, with the interpretation that s_1 must be more similar to s_2 than to s_3, formulated as follows: $\forall (s_1, s_2, s_3) \in \text{trip}_{\mathbf{s}}, \forall z_1, z_2, z_3 \in \mathcal{Z}, \quad \forall i, j, k \in [k]$, if $s_1, z_1 \in C_i, s_2, z_2 \in C_j$ and $s_3, z_3 \in C_k$ then

$$|\ell(\mathcal{A}_{\text{trip}_{\mathbf{s}}}, s_1, s_2, s_3) - \ell(\mathcal{A}_{\text{trips}_{\mathbf{s}}}, z_1, z_2, z_3)| \leq \epsilon(\text{trip}_{\mathbf{s}}). \tag{A.1}$$

Following this definition, Proposition 3 can be easily extended to obtain the following generalization bound

$$\left|\mathcal{L}(\mathcal{A}_{\text{trip}_{\mathbf{s}}}) - \mathcal{L}_{\text{emp}}(\mathcal{A}_{\text{trip}_{\mathbf{s}}})\right| \leq \epsilon(\text{trip}_{\mathbf{s}}) + 3B\sqrt{\frac{2K \ln 2 + 2 \ln 1/\delta}{n}}. \tag{A.2}$$

For triplet based metric learning algorithms, by following the definition of robustness given by Eq. (A.1) and adapting straightforwardly the losses to triplets such that they output zero for non-admissible ones, we can obtain the following bound

$$|\mathcal{L}(\mathcal{A}_{\mathbf{s}}) - \mathcal{L}_{\mathbf{emp}}(\mathcal{A}_{\mathbf{s}})| \leq \epsilon(\mathbf{s}) + \xi(\mathcal{A}_{\mathbf{s}})\left((3\sqrt{2}+3)\sqrt{\frac{|\mathcal{T}_{\mathbf{s}}| \ln(2K/\delta)}{n}} + \frac{2|\mathcal{T}_{\mathbf{s}}| \ln(2K/\delta)}{n}\right). \tag{A.3}$$

Using triplet-based robustness, consider algorithms of the following form

$$\min_{\mathbf{M} \succeq 0} \; c\|\mathbf{M}\| + \frac{1}{|\text{trip}_s|} \sum_{(s_i, s_j, s_k) \in \text{trip}_s} \left[1 - (x_i - x_k)^T \mathbf{M}(x_i - x_k) + (x_i - x_j)^T \mathbf{M}(x_i - x_j)\right]_+,$$

where $\|\mathbf{M}\| = \|\mathbf{M}\|_{\mathcal{F}}$ in Example 4 or $\|\mathbf{M}\| = \|\mathbf{M}\|_{1,2}$ in Example 5. These methods are $(\mathcal{N}(\gamma, \mathbb{Z}, \|\cdot\|_2), 16UR_y g_0/c)$-robust.

Appendix B. Additional Proofs

In this section, we present the proof of Lemma 17, Lemma 18 and Lemma 19. Recall that C_i is defined in Definition 1.

Proof [Proof of Lemma 17] By the definition of $\mathcal{L}$ and $\mathcal{L}_{\mathbf{emp}}$, we know

$$
\begin{aligned}
&|\mathcal{L}\left(\mathcal{A}_{\mathbf{s}}\right) - \mathcal{L}_{\mathbf{emp}}\left(\mathcal{A}_{\mathbf{s}}\right)| \\
&= \left| \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z')] - \frac{1}{n^2} \sum_{i,j=1}^{n} \ell(\mathcal{A}_{\mathbf{s}}, z_i, z_j) \right| \\
&= \left| \sum_{i,j \in [K]} \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') | z \in C_i, z' \in C_j] p_i p_j - \frac{1}{n^2} \sum_{i,j=1}^{n} \ell(\mathcal{A}_{\mathbf{s}}, z_i, z_j) \right| \\
&\overset{(a)}{\leq} \left| \sum_{i,j \in [K]} \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') | z \in C_i, z' \in C_j] p_i p_j - \sum_{i,j \in [K]} \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') | z \in C_i, z' \in C_j] p_i \frac{|\mathcal{I}_j|}{n} \right| \\
&+ \left| \sum_{i,j \in [K]} \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') | z \in C_i, z' \in C_j] p_i \frac{|\mathcal{I}_j|}{n} - \frac{1}{n^2} \sum_{i,j=1}^{n} \ell(\mathcal{A}_{\mathbf{s}}, z_i, z_j) \right| \\
&\overset{(b)}{\leq} \left| \sum_{i,j \in [K]} \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') | z \in C_i, z' \in C_j] p_i (p_j - \frac{|\mathcal{I}_j|}{n}) \right| \\
&+ \left| \sum_{i,j \in [K]} \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') | z \in C_i, z' \in C_j] p_i \frac{|\mathcal{I}_j|}{n} - \sum_{i,j \in [K]} \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') | z \in C_i, z' \in C_j] \frac{|\mathcal{I}_i||\mathcal{I}_j|}{n^2} \right| \\
&+ \left| \sum_{i,j \in [K]} \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') | z \in C_i, z' \in C_j] \frac{|\mathcal{I}_i||\mathcal{I}_j|}{n^2} - \frac{1}{n^2} \sum_{i,j=1}^{n} \ell(\mathcal{A}_{\mathbf{s}}, z_i, z_j) \right|,
\end{aligned}
$$

where inequalities (a) and (b) are due to the triangle inequality. By the following symmetrical property of metric $\mathcal{A}_{\mathbf{s}}$, we know

$$
\begin{aligned}
&\left| \sum_{i,j \in [K]} \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') | z \in C_i, z' \in C_j] \frac{|\mathcal{I}_j|}{n} (p_i - \frac{|\mathcal{I}_i|}{n}) \right| \\
=&\left| \sum_{i,j \in [K]} \mathbb{E}_{z,z' \sim \mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') | z \in C_i, z' \in C_j] \frac{|\mathcal{I}_i|}{n} (p_j - \frac{|\mathcal{I}_j|}{n}) \right|.
\end{aligned}
$$

It then follows that

$$
\begin{aligned}
&|\mathcal{L}\left(\mathcal{A}_{\mathbf{s}}\right)-\mathcal{L}_{\mathbf{emp}}\left(\mathcal{A}_{\mathbf{s}}\right)| \\
\leq&\left|\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j]p_i(p_j-\frac{|\mathcal{I}_j|}{n})\right| \\
+&\left|\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j]\frac{|\mathcal{I}_j|}{n}(p_i-\frac{|\mathcal{I}_i|}{n})\right| \\
+&\left|\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j]\frac{|\mathcal{I}_i||\mathcal{I}_j|}{n^2}-\frac{1}{n^2}\sum_{i,j=1}^{n}\ell(\mathcal{A}_{\mathbf{s}},z_i,z_j)\right| \\
\leq&\left|\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j]p_j(p_i-\frac{|\mathcal{I}_i|}{n})\right| \\
+&\left|\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j]\frac{|\mathcal{I}_j|}{n}(p_i-\frac{|\mathcal{I}_i|}{n})\right| \\
+&\left|\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j]\frac{|\mathcal{I}_i||\mathcal{I}_j|}{n^2}-\frac{1}{n^2}\sum_{i,j=1}^{n}\ell(\mathcal{A}_{\mathbf{s}},z_i,z_j)\right| \\
\leq&\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j](p_j+\frac{|\mathcal{I}_j|}{n})\left|p_i-\frac{|\mathcal{I}_i|}{n}\right| \\
+&\left|\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j]\frac{|\mathcal{I}_i||\mathcal{I}_j|}{n^2}-\frac{1}{n^2}\sum_{i,j=1}^{n}\ell(\mathcal{A}_{\mathbf{s}},z_i,z_j)\right|.
\end{aligned}
$$

The proof is completed. ■

Proof [Proof of Lemma 18] By definition, we have $\sum_{i=1}^{K}|\mathcal{I}_i|=n$ and $\sum_{i=1}^{K}p_i=1$. Then,

$$
\begin{aligned}
&\left|\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j]\frac{|\mathcal{I}_i||\mathcal{I}_j|}{n^2}-\frac{1}{n^2}\sum_{i,j=1}^{n}\ell(\mathcal{A}_{\mathbf{s}},z_i,z_j)\right| \\
\leq&\frac{1}{n^2}\left|\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j]|\mathcal{I}_i||\mathcal{I}_j|-\sum_{i,j=1}^{n}\ell(\mathcal{A}_{\mathbf{s}},z_i,z_j)\right| \\
=&\frac{1}{n^2}\left|\sum_{i,j\in[K]}\mathbb{E}_{z,z'\sim\mu}\sum_{z_p\in C_i,z_q\in C_j}[\ell(\mathcal{A}_{\mathbf{s}},z,z')|z\in C_i,z'\in C_j]-\sum_{i,j\in[K]}\sum_{z_p\in C_i,z_q\in C_j}\ell(\mathcal{A}_{\mathbf{s}},z_p,z_q)\right| \\
\leq&\frac{1}{n^2}\sum_{i,j\in[K]}\sum_{z_p\in C_i,z_q\in C_j}\max_{z\in C_i,z'\in C_j}\left|\ell(\mathcal{A}_{\mathbf{s}},z,z')-\ell(\mathcal{A}_{\mathbf{s}},z_p,z_q)\right| \\
\leq&\max_{i,j\in[n]}\max_{z_p,z\in C_i,z_q,z'\in C_j}\left|\ell(\mathcal{A}_{\mathbf{s}},z,z')-\ell(\mathcal{A}_{\mathbf{s}},z_p,z_q)\right|.
\end{aligned}
$$

■

Proof [Proof of Lemma 19] First, define $\hat{\mathcal{I}}_k := \{i \in [n] : z_i \in \hat{\mathbf{s}}, z_i \in C_k\}$, $\hat{\mathbf{s}}$ is defined in definition 12. Then starting from the second to the last step of Lemma 18, we know

$$\begin{aligned}
&\left| \sum_{i,j\in[K]} \mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z')|z \in C_i, z' \in C_j] \frac{|\mathcal{I}_i||\mathcal{I}_j|}{n^2} - \frac{1}{n^2}\sum_{i,j=1}^{n} \ell(\mathcal{A}_{\mathbf{s}}, z_i, z_j) \right| \\
&\leq \frac{1}{n^2} \sum_{i,j\in[K]} \sum_{z_p\in C_i, z_q\in C_j} \max_{z\in C_i, z'\in C_j} \left|\ell(\mathcal{A}_{\mathbf{s}}, z, z') - \ell(\mathcal{A}_{\mathbf{s}}, z_p, z_q)\right| \\
&\leq \frac{1}{n^2} \sum_{i,j\in[K]} \sum_{p\in\hat{\mathcal{I}}_i, q\in\hat{\mathcal{I}}_j} \max_{z\in C_i, z'\in C_j} \left|\ell(\mathcal{A}_{\mathbf{s}}, z, z') - \ell(\mathcal{A}_{\mathbf{s}}, z_p, z_q)\right| \\
&+ \frac{1}{n^2} \sum_{i,j\in[K]} \sum_{\neg(p\in\hat{\mathcal{I}}_i, q\in\hat{\mathcal{I}}_j)} \max_{z\in C_i, z'\in C_j} \left|\ell(\mathcal{A}_{\mathbf{s}}, z, z') - \ell(\mathcal{A}_{\mathbf{s}}, z_p, z_q)\right| \\
&\leq \frac{1}{n^2} \sum_{i,j\in[K]} \sum_{p\in\hat{\mathcal{I}}_i, q\in\hat{\mathcal{I}}_j} \max_{z\in C_i, z'\in C_j} \left|\ell(\mathcal{A}_{\mathbf{s}}, z, z') - \ell(\mathcal{A}_{\mathbf{s}}, z_p, z_q)\right| \\
&+ \frac{1}{n^2} \sum_{i,j\in[K]} \sum_{\neg(p\in\hat{\mathcal{I}}_i, q\in\hat{\mathcal{I}}_j)} \max_{z\in C_i, z'\in C_j} \left|\ell(\mathcal{A}_{\mathbf{s}}, z, z') - \ell(\mathcal{A}_{\mathbf{s}}, z_p, z_q)\right|.
\end{aligned}$$

Recall that ℓ is positive, and the weak robustness properties. It then follows that

$$\begin{aligned}
&\left| \sum_{i,j\in[K]} \mathbb{E}_{z,z'\sim\mu}[\ell(\mathcal{A}_{\mathbf{s}}, z, z')|z \in C_i, z' \in C_j] \frac{|\mathcal{I}_i||\mathcal{I}_j|}{n^2} - \frac{1}{n^2}\sum_{i,j=1}^{n} \ell(\mathcal{A}_{\mathbf{s}}, z_i, z_j) \right| \\
&\leq \frac{|\hat{\mathbf{s}}^2|}{n^2} + \frac{n^2 - |\hat{\mathbf{s}}^2|}{n^2}\xi\left(\mathcal{A}_{\mathbf{s}}\right).
\end{aligned}$$

The proof is completed. ■

Appendix C. Notation Index

$\mathcal{Z}$ Sample space, defined as $\mathcal{X} \times \mathcal{Y}$.

$\mathcal{X}$ Input space, a subset of $\mathbb{R}^d$, where d is the dimensionality of the input space.

$\mathcal{Y}$ Output space, a subset of $\mathbb{R}$.

μ Unknown probability distribution over $\mathcal{Z}$.

$\mathbf{s}$ Training dataset, $\mathbf{s} = (z_1, \ldots, z_n)$.

n Number of training examples.

z_i Individual training example, $z_i \in \mathcal{Z}$.

$\mathcal{F}$ Model space.

f Model in model space $\mathcal{F}$; $f : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ in pairwise learning or $f : \mathcal{X} \mapsto \mathbb{R}$ in pointwise learning.

ℓ Loss function; $\ell : \mathcal{F} \times \mathcal{Z} \times \mathcal{Z} \mapsto \mathbb{R}^+$.

$\ell(f, z, z')$ Loss of model f on example pair (z, z') with $\ell(f, z, z) = 0$.

$\mathcal{L}(f)$ Population risk: $\mathbb{E}_{z,z' \sim \mu}[\ell(f, z, z')]$.

$\mathcal{L}_{\mathbf{emp}}(f)$ Empirical risk; $\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \ell(f, z_i, z_j)$.

$\mathcal{A}$ Optimization algorithm.

$\mathcal{A}_{\mathbf{s}}$ Model learned by applying $\mathcal{A}$ to dataset $\mathbf{s}$.

B Upper bound on loss: $\ell(f, z) \leq B$ for all $f \in \mathcal{F}, z \in \mathcal{Z}$.

$|\mathcal{B}|$ Cardinality (number of elements) of the set $\mathcal{B}$

$[n]$ Set of integers $\{1, \ldots, n\}$ where $n \in \mathbb{N}$

$\epsilon(\cdot)$ Robustness parameter function $\epsilon : \mathcal{Z}^n \mapsto \mathbb{R}_+$

K Number of partitions of Sample space $\mathcal{Z}$.

$\{C_k\}_{k=1}^{K}$ Partition of $\mathcal{Z}$ into K disjoint sets

$\mathcal{T}_{\mathbf{s}}$ Set of partition indices with at least one training example: $\{k \in [K] : |\mathcal{I}_k^{\mathbf{s}}| \geq 1\}$

$\mathcal{T}_{\mathbf{s}}^c$ Complement of $\mathcal{T}_{\mathbf{s}}$: $\{k \in [K] : |\mathcal{I}_k^{\mathbf{s}}| = 0\}$ (partitions with no training examples)

$\mathcal{I}_k^{\mathbf{s}}$ Index set of examples in $\mathbf{s}$ belonging to C_k: $\{i \in [n] : z_i \in C_k\}$

p_k Probability $\mathbb{P}(z \in C_k)$

p Probability vector $(p_1, \ldots, p_K)$

$\xi(\mathcal{A}_{\mathbf{s}})$ Conditional expected loss: $\max_{i,j \in [K]} \mathbb{E}_{z,z'}[\ell(\mathcal{A}_{\mathbf{s}}, z, z') \mid z \in C_i, z' \in C_j]$

δ Confidence parameter ($\delta \in (0, 1)$)

$\alpha_k(f)$ Maximum conditional expected loss for class k: $\max_{j \in [n]} \mathbb{E}_{z,z'}[\ell(f, z, z') \mid z \in C_k, z' \in C_j]$

$\alpha_{\mathcal{T}_{\mathbf{s}}}(f)$ Maximum α_k over active partitions: $\max_{k \in \mathcal{T}_{\mathbf{s}}} \alpha_k(f)$

$\alpha_{\mathcal{T}_{\mathbf{s}}^c}(f)$ Maximum α_k over inactive partitions: $\max_{k \in \mathcal{T}_{\mathbf{s}}^c} \alpha_k(f)$

$\mathcal{Q}_1 = \sum_{k \in \mathcal{T}_{\mathbf{s}}} \left(\alpha_{\mathcal{T}_{\mathbf{s}}^c}(\mathcal{A}_{\mathbf{s}}) + \sqrt{2} \alpha_k(\mathcal{A}_{\mathbf{s}}) \sqrt{\frac{|\mathcal{I}_k^{\mathbf{s}}|}{n}} \right)$

$\mathcal{Q}_2 = \alpha_{\mathcal{T}_{\mathbf{s}}^c}(\mathcal{A}_{\mathbf{s}}) |\mathcal{T}_{\mathbf{s}}| + \sum_{k \in \mathcal{T}_{\mathbf{s}}} \alpha_k(\mathcal{A}_{\mathbf{s}})$

$\hat{p}_n(\cdot)$ Pseudo-robustness function: $\hat{p}_n : (\mathcal{Z} \times \mathcal{Z})^n \to \{1, \ldots, n^2\}$

$\mathbf{s}^2$ Full set of training pairs: $\{(z_i, z_j) : z_i, z_j \in \mathbf{s}\}$ (size n^2)

$\hat{\mathbf{s}}^2$ Subset of pairwise training dataset $\mathbf{s}^2$.

$\epsilon(\mathbf{s}^2)$ Robustness parameter as function of training pairs