

A Proof of Theorem 2

From Eq.(11), we have

$$\begin{aligned}\sum_{\bar{y}=1}^K \bar{\mathcal{L}}_{\text{OVA}}(f(\mathbf{x}), \bar{y}) &= \frac{1}{K-1} \sum_{\bar{y}=1}^K \sum_{y \neq \bar{y}} \ell(g_y(\mathbf{x})) + \sum_{\bar{y}=1}^K \ell(-g_{\bar{y}}(\mathbf{x})) \\ &= \sum_{\bar{y}=1}^K (\ell(g_{\bar{y}}(\mathbf{x})) + \ell(-g_{\bar{y}}(\mathbf{x}))) = K.\end{aligned}$$

$$\begin{aligned}\mathcal{L}_{\text{OVA}}(f(\mathbf{x}), y) + \bar{\mathcal{L}}_{\text{OVA}}(f(\mathbf{x}), y) &= \ell(g_y(\mathbf{x})) + \frac{1}{K-1} \sum_{\bar{y} \neq y} \ell(-g_{\bar{y}}(\mathbf{x})) \\ &\quad + \frac{1}{K-1} \sum_{y' \neq y} \ell(g_{y'}(\mathbf{x})) + \ell(-g_y(\mathbf{x})) = 2,\end{aligned}$$

$$\begin{aligned}\sum_{\bar{y}=1}^K \bar{\mathcal{L}}_{\text{PC}}(f(\mathbf{x}), \bar{y}) &= \sum_{\bar{y}=1}^K \sum_{y \neq \bar{y}} \ell(g_y(\mathbf{x}) - g_{\bar{y}}(\mathbf{x})) \\ &= \sum_{\bar{y}=1}^{K-1} \sum_{y=\bar{y}+1}^K (\ell(g_y(\mathbf{x}) - g_{\bar{y}}(\mathbf{x})) + \ell(g_{\bar{y}}(\mathbf{x}) - g_y(\mathbf{x}))) = \frac{K(K-1)}{2},\end{aligned}$$

$$\mathcal{L}_{\text{PC}}(f(\mathbf{x}), y) + \bar{\mathcal{L}}_{\text{PC}}(f(\mathbf{x}), y) = \sum_{y' \neq y} \ell(g_y(\mathbf{x}) - g_{y'}(\mathbf{x})) + \sum_{y' \neq y} \ell(g_{y'}(\mathbf{x}) - g_y(\mathbf{x})) = K-1.$$

□

B Proof of Lemma 3

By definition, $h(\mathbf{x}_i, \bar{y}_i) = \tilde{\mathcal{L}}_{\text{OVA}}(f(\mathbf{x}_i), \bar{y}_i)$ so that

$$\bar{\mathfrak{R}}_n(\mathcal{H}_{\text{OVA}}) = \mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \left(\frac{1}{K-1} \sum_{y \neq \bar{y}} \tilde{\ell}(g_y(\mathbf{x}_i)) + \tilde{\ell}(-g_{\bar{y}_i}(\mathbf{x}_i)) \right) \right].$$

After rewriting $\tilde{\mathcal{L}}_{\text{OVA}}(f(\mathbf{x}_i), \bar{y}_i)$, we can know that

$$\tilde{\mathcal{L}}_{\text{OVA}}(f(\mathbf{x}_i), \bar{y}_i) = \frac{1}{K-1} \sum_y \tilde{\ell}(g_y(\mathbf{x}_i)) + \frac{K-2}{K-1} \tilde{\ell}(-g_{\bar{y}_i}(\mathbf{x}_i)),$$

and subsequently,

$$\begin{aligned}\bar{\mathfrak{R}}_n(\mathcal{H}_{\text{OVA}}) &\leq \frac{1}{K-1} \mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \sum_y \tilde{\ell}(g_y(\mathbf{x}_i)) \right] \\ &\quad + \frac{K-2}{K-1} \mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \tilde{\ell}(-g_{\bar{y}_i}(\mathbf{x}_i)) \right]\end{aligned}$$

due to the sub-additivity of the supremum.

The first term is independent of \bar{y}_i and thus

$$\begin{aligned}
\mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \sum_y \tilde{\ell}(g_y(\mathbf{x}_i)) \right] &= \mathbb{E}_{\bar{\mathcal{X}}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{\mathbf{x}_i \in \bar{\mathcal{X}}} \sigma_i \sum_y \tilde{\ell}(g_y(\mathbf{x}_i)) \right] \\
&\leq \sum_y \mathbb{E}_{\bar{\mathcal{X}}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{\mathbf{x}_i \in \bar{\mathcal{X}}} \sigma_i \tilde{\ell}(g_y(\mathbf{x}_i)) \right] \\
&= \sum_y \mathbb{E}_{\bar{\mathcal{X}}} \mathbb{E}_{\sigma} \left[\sup_{g_y \in \mathcal{G}} \frac{1}{n} \sum_{\mathbf{x}_i \in \bar{\mathcal{X}}} \sigma_i \tilde{\ell}(g_y(\mathbf{x}_i)) \right] \\
&= K \bar{\mathfrak{R}}_n(\tilde{\ell} \circ \mathcal{G})
\end{aligned}$$

which means the first term can be bounded by $K/(K-1) \cdot \bar{\mathfrak{R}}_n(\tilde{\ell} \circ \mathcal{G})$. The second term is more involved. Let $I(\cdot)$ be the indicator function and $\alpha_i = 2I(y = \bar{y}_i) - 1$, then

$$\begin{aligned}
&\mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \tilde{\ell}(-g_{\bar{y}_i}(\mathbf{x}_i)) \right] \\
&= \mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \sum_y \tilde{\ell}(-g_y(\mathbf{x}_i)) I(y = \bar{y}_i) \right] \\
&= \mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{2n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \sum_y \tilde{\ell}(-g_y(\mathbf{x}_i)) (\alpha_i + 1) \right] \\
&\leq \mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{2n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \alpha_i \sigma_i \sum_y \tilde{\ell}(-g_y(\mathbf{x}_i)) \right] \\
&\quad + \mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{2n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \sum_y \tilde{\ell}(-g_y(\mathbf{x}_i)) \right] \\
&= \mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \sum_y \tilde{\ell}(-g_y(\mathbf{x}_i)) \right],
\end{aligned}$$

where we used that $\alpha_i \sigma_i$ has exactly the same distribution as σ_i . This can be similarly bounded by $\bar{\mathfrak{R}}_n(\tilde{\ell} \circ \mathcal{G})$ and the second term can be bounded by $K(K-2)/(K-1) \cdot \bar{\mathfrak{R}}_n(\tilde{\ell} \circ \mathcal{G})$.

As a result,

$$\begin{aligned}
\bar{\mathfrak{R}}_n(\mathcal{H}_{\text{OVA}}) &\leq \left(\frac{K}{K-1} + \frac{K(K-2)}{K-1} \right) \bar{\mathfrak{R}}_n(\tilde{\ell} \circ \mathcal{G}) \\
&= K \bar{\mathfrak{R}}_n(\tilde{\ell} \circ \mathcal{G}) \\
&\leq K L_{\ell} \bar{\mathfrak{R}}_n(\mathcal{G}) \\
&= K L_{\ell} \mathfrak{R}_n(\mathcal{G}),
\end{aligned}$$

according to Talagrand's contraction lemma [19]. \square

C Proof of Lemma 4

By definition,

$$\bar{\mathfrak{R}}_n(\mathcal{H}_{\text{PC}}) = \mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \left(\sum_{y' \neq \bar{y}_i} \tilde{\ell}(g_{y'}(\mathbf{x}_i) - g_{\bar{y}_i}(\mathbf{x}_i)) \right) \right].$$

Using the proof technique for handling the second term in the proof of Lemma 3, we have

$$\begin{aligned}
\overline{\mathfrak{R}}_n(\mathcal{H}_{\text{PC}}) &\leq \mathbb{E}_{\mathcal{S}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{(\mathbf{x}_i, \bar{y}_i) \in \mathcal{S}} \sigma_i \sum_y \left(\sum_{y' \neq y} \tilde{\ell}(g_{y'}(\mathbf{x}_i) - g_y(\mathbf{x}_i)) \right) \right] \\
&= \mathbb{E}_{\bar{\mathcal{X}}} \mathbb{E}_{\sigma} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} \frac{1}{n} \sum_{\mathbf{x}_i \in \bar{\mathcal{X}}} \sigma_i \sum_y \left(\sum_{y' \neq y} \tilde{\ell}(g_{y'}(\mathbf{x}_i) - g_y(\mathbf{x}_i)) \right) \right] \\
&\leq \sum_y \sum_{y' \neq y} \mathbb{E}_{\bar{\mathcal{X}}} \mathbb{E}_{\sigma} \left[\sup_{g_y, g_{y'} \in \mathcal{G}} \frac{1}{n} \sum_{\mathbf{x}_i \in \bar{\mathcal{X}}} \sigma_i \tilde{\ell}(g_{y'}(\mathbf{x}_i) - g_y(\mathbf{x}_i)) \right],
\end{aligned}$$

due to the sub-additivity of the supremum.

Let

$$\mathcal{G}_{y, y'} = \{\mathbf{x} \mapsto g_{y'}(\mathbf{x}) - g_y(\mathbf{x}) \mid g_y, g_{y'} \in \mathcal{G}\},$$

then according to Talagrand's contraction lemma [19],

$$\begin{aligned}
&\mathbb{E}_{\bar{\mathcal{X}}} \mathbb{E}_{\sigma} \left[\sup_{g_y, g_{y'} \in \mathcal{G}} \frac{1}{n} \sum_{\mathbf{x}_i \in \bar{\mathcal{X}}} \sigma_i \tilde{\ell}(g_{y'}(\mathbf{x}_i) - g_y(\mathbf{x}_i)) \right] \\
&= \overline{\mathfrak{R}}_n(\tilde{\ell} \circ \mathcal{G}_{y, y'}) \\
&\leq L_{\ell} \overline{\mathfrak{R}}_n(\mathcal{G}_{y, y'}) \\
&= L_{\ell} \mathbb{E}_{\bar{\mathcal{X}}} \mathbb{E}_{\sigma} \left[\sup_{g_y, g_{y'} \in \mathcal{G}} \frac{1}{n} \sum_{\mathbf{x}_i \in \bar{\mathcal{X}}} \sigma_i (g_{y'}(\mathbf{x}_i) - g_y(\mathbf{x}_i)) \right] \\
&\leq L_{\ell} \mathbb{E}_{\bar{\mathcal{X}}} \mathbb{E}_{\sigma} \left[\sup_{g_y \in \mathcal{G}} \frac{1}{n} \sum_{\mathbf{x}_i \in \bar{\mathcal{X}}} \sigma_i g_y(\mathbf{x}_i) \right] + L_{\ell} \mathbb{E}_{\bar{\mathcal{X}}} \mathbb{E}_{\sigma} \left[\sup_{g_{y'} \in \mathcal{G}} \frac{1}{n} \sum_{\mathbf{x}_i \in \bar{\mathcal{X}}} \sigma_i g_{y'}(\mathbf{x}_i) \right] \\
&= 2L_{\ell} \overline{\mathfrak{R}}_n(\mathcal{G}) \\
&= 2L_{\ell} \mathfrak{R}_n(\mathcal{G}).
\end{aligned}$$

This proves that $\overline{\mathfrak{R}}_n(\mathcal{H}_{\text{PC}}) \leq 2K(K-1)L_{\ell}\mathfrak{R}_n(\mathcal{G})$. \square

D Proof of Lemma 5

We are going to prove the case of $\bar{\mathcal{L}}_{\text{OVA}}$; the other case is similar. We consider a single direction $\sup_{g_1, \dots, g_K \in \mathcal{G}} (\hat{R}(f) - R(f))$ with probability at least $1 - \delta/2$; the other direction is similar too.

Given the symmetric condition (11), it must hold that $\|\bar{\mathcal{L}}_{\text{OVA}}\|_{\infty} = 2$ when g_1, \dots, g_K can be any measurable functions. Let a single $(\mathbf{x}_i, \bar{y}_i)$ be replaced with $(\mathbf{x}'_i, \bar{y}'_i)$, then the change of $\sup_{g_1, \dots, g_K \in \mathcal{G}} (\hat{R}(f) - R(f))$ is no greater than $2(K-1)/n$. Apply McDiarmid's inequality [22] to the single-direction uniform deviation $\sup_{g_1, \dots, g_K \in \mathcal{G}} (\hat{R}(f) - R(f))$ to get that

$$\Pr \left\{ \sup_{g_1, \dots, g_K \in \mathcal{G}} (\hat{R}(f) - R(f)) - \mathbb{E} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} (\hat{R}(f) - R(f)) \right] \geq \epsilon \right\} \leq \exp \left(-\frac{2\epsilon^2}{n(2(K-1)/n)^2} \right),$$

or equivalently, with probability at least $1 - \delta/2$,

$$\sup_{g_1, \dots, g_K \in \mathcal{G}} (\hat{R}(f) - R(f)) \leq \mathbb{E} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} (\hat{R}(f) - R(f)) \right] + (K-1) \sqrt{\frac{2 \ln(2/\delta)}{n}}.$$

Since $R(f) = \mathbb{E}[\hat{R}(f)]$, it is a routine work to show by symmetrization that [23]

$$\begin{aligned}
\mathbb{E} \left[\sup_{g_1, \dots, g_K \in \mathcal{G}} (\hat{R}(f) - R(f)) \right] &\leq 2(K-1) \overline{\mathfrak{R}}_n(\mathcal{H}_{\text{OVA}}) \\
&\leq 2K(K-1)L_{\ell}\mathfrak{R}_n(\mathcal{G}),
\end{aligned}$$

where the last line is due to Lemma 3. \square