# LEVERAGED WEIGHTED LOSS FOR PARTIAL LABEL LEARNING

**Anonymous authors**
Paper under double-blind review

## 1 PROOFS

We present all proofs for Section 2 here. For the sake of conciseness and readability, we denote $\boldsymbol{\mathcal{Y}}^y$ as the collection of all partial label sets containing the true label $y$, i.e. $\boldsymbol{\mathcal{Y}}^y := \{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}} | y \in \boldsymbol{y}\}$.

**Proof 1 (of Lemma 1)** *When Assumption 3 holds, elementary probability theory states that*

$$\begin{aligned}\mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} | Y = y, X = x) &= \prod_{s \in \boldsymbol{y}} q_s \cdot \prod_{t \notin \boldsymbol{y}} (1 - q_t) \\ &= q_y \prod_{s \in \boldsymbol{y}, s \neq y} q_s \cdot \prod_{t \notin \boldsymbol{y}} (1 - q_t).\end{aligned}$$

*By Assumption 1, we have $q_z = 1$, and thus complete the proof.*

**Proof 2 (of Theorem 1)** *For any $x \in \mathcal{X}$, there holds*

$$\begin{aligned}\bar{\mathcal{R}}(\bar{\mathcal{L}}, g(X)) &= \mathbb{E}_{\boldsymbol{Y}|X}[\bar{\mathcal{L}}(\boldsymbol{Y}, g(x)) | X = x] \\ &= \sum_{\boldsymbol{y} \in 2^{[K]}} \bar{\mathcal{L}}(\boldsymbol{y}, g(x)) \mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} | X = x) \\ &= \sum_{\boldsymbol{y} \in 2^{[K]}} \bar{\mathcal{L}}(\boldsymbol{y}, g(x)) \sum_{y \in \boldsymbol{y}} \mathrm{P}(\boldsymbol{Y} = \boldsymbol{y}, Y = y | X = x) \\ &= \sum_{\boldsymbol{y} \in 2^{[K]}} \bar{\mathcal{L}}(\boldsymbol{y}, g(x)) \sum_{y \in \boldsymbol{y}} \mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} | Y = y, X = x) \mathrm{P}(Y = y | X = x) \\ &= \sum_{y=1}^{K} \sum_{\boldsymbol{y} \in 2^{[K]}} \mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} | Y = y, X = x) \bar{\mathcal{L}}(\boldsymbol{y}, g(x)) \mathrm{P}(Y = y | X = x),\end{aligned}$$

*and*

$$\mathcal{R}(\mathcal{L}, g(X)) = \mathbb{E}_{Y|X}[\mathcal{L}(Y, g(x)) | X = x] = \sum_{y=1}^{K} \mathcal{L}(y, g(x)) \mathrm{P}(Y = y | X = x).$$

*Therefore, if we have*

$$\begin{aligned}\mathcal{L}(y, g(x)) &= \sum_{\boldsymbol{y} \in 2^{[K]}} \mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} | Y = y, X = x) \bar{\mathcal{L}}(\boldsymbol{y}, g(x)) \\ &= \sum_{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}^y} \mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} | Y = y, X = x) \bar{\mathcal{L}}(\boldsymbol{y}, g(x)),\end{aligned}$$

*since $\mathrm{P}(\boldsymbol{Y} = \boldsymbol{y} | Y = y, X = x) = 0$ for $\boldsymbol{y}$ not containing $y$, then there holds*

$$\bar{\mathcal{R}}(\bar{\mathcal{L}}, g(X)) = \mathcal{R}(\mathcal{L}, g(X)).$$

*According to Lemma 1, by Assumption 1 and Assumption 3, there holds*

$$\mathcal{L}(y, g(x)) = \sum_{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}^y} \prod_{s \in \boldsymbol{y}, s \neq y} q_s \prod_{t \notin \boldsymbol{y}} (1 - q_t) \bar{\mathcal{L}}(\boldsymbol{y}, g(x)).$$

**Lemma 1** *Let $y$ be the true label of input $x$, $q_z := \mathrm{P}(z \in \boldsymbol{y}|Y = y, X = x)$ for $z \in \mathcal{Y}$, and $\boldsymbol{\mathcal{Y}}^y := \{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}|y \in \boldsymbol{y}\}$. Then there holds*

$$\sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t) = 1.$$

**Proof 3 (of Lemma 1)** *By Assumption 1, we have*

$$\begin{aligned}\sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t) &= \sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} 1\cdot q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)\\ &= \sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_y\cdot q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)\\ &= \sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y}} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)\\ &= \sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \mathrm{P}(\boldsymbol{Y}=\boldsymbol{y}|Y=y,X=x)\\ &= 1,\end{aligned}$$

*where the last equation holds since $\mathrm{P}(\boldsymbol{Y} = \boldsymbol{y}|Y = y, X = x) = 0$ for $\boldsymbol{y} \notin \boldsymbol{\mathcal{Y}}^y$.*

**Proof 4 (of Theorem 2)** *According to Theorem 1, we have the form of the corresponding ordinary loss function being*

$$\begin{aligned}\mathcal{L}_\psi(y,g(x)) &= \sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)\bar{\mathcal{L}}_\psi(\boldsymbol{y},g(x))\\ &= \sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)\sum_{z\in\boldsymbol{y}} w_z\psi(g_z(x))\\ &+ \beta\cdot\sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)\sum_{z\notin\boldsymbol{y}} w_z\psi(-g_z(x)).\end{aligned} \tag{1}$$

*The first term on the right hand side of (1) is*

$$\begin{aligned}&\sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)\sum_{z\in\boldsymbol{y}} w_z\psi(g_z(x))\\ =&\sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)w_y\psi(g_y(x)) + \sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)\sum_{z\in\boldsymbol{y}\setminus\{y\}} w_z\psi(g_z(x))\\ =&\sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)w_y\psi(g_y(x)) + \sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y}\sum_{z\in\boldsymbol{y}\setminus\{y\}} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\in[K]\setminus\boldsymbol{y}}(1-q_t)w_z\psi(g_z(x)).\end{aligned} \tag{2}$$

*By Lemma 1, we have*

$$\sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t) = 1,$$

*and therefore the first term in (2) becomes*

$$\sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\notin\boldsymbol{y}}(1-q_t)w_y\psi(g_y(x)) = w_y\psi(g_y(x)). \tag{3}$$

*For the second term in (2), since $z \neq y$ and $z \in \boldsymbol{y}$, we switch the summations, and achieve*

$$\begin{aligned}&\sum_{\boldsymbol{y}\in 2^{[K]}}\sum_{z\in\boldsymbol{y}\setminus\{y\}} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\in[K]\setminus\boldsymbol{y}}(1-q_t)w_z\psi(g_z(x))\\ =&\sum_{z\neq y}\sum_{\boldsymbol{y}\in\boldsymbol{\mathcal{Y}}^z\cap\boldsymbol{\mathcal{Y}}^y} \prod_{s\in\boldsymbol{y},s\neq y} q_s \prod_{t\in[K]\setminus\boldsymbol{y}}(1-q_t)w_z\psi(g_z(x))\end{aligned}$$

$$= \sum_{z \neq y} w_z \psi(g_z(x)) \sum_{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}^z \setminus \boldsymbol{\mathcal{Y}}^y} \prod_{s \in \boldsymbol{y}} q_s \prod_{t \in [K] \setminus \boldsymbol{y} \setminus \{y\}} (1 - q_t).$$

*Without loss of generality, we assume $y = K$ for notational simplicity, and write*

$$\begin{aligned}
&\sum_{z \neq y} w_z \psi(g_z(x)) \sum_{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}^z \setminus \boldsymbol{\mathcal{Y}}^y} \prod_{s \in \boldsymbol{y}} q_s \prod_{t \in [K] \setminus \boldsymbol{y} \setminus \{y\}} (1 - q_t) \\
&= \sum_{z \in [K-1]} w_z \psi(g_z(x)) \sum_{\boldsymbol{y} \in (2^{[K-1]})^z} \prod_{s \in \boldsymbol{y}} q_s \prod_{t \in [K-1] \setminus \boldsymbol{y}} (1 - q_t) \\
&= \sum_{z \in [K-1]} w_z \psi(g_z(x)) q_z \sum_{\boldsymbol{y} \in (2^{[K-1]})^z} \prod_{s \in \boldsymbol{y}, s \neq z} q_s \prod_{t \in [K-1] \setminus \boldsymbol{y}} (1 - q_t).
\end{aligned}$$

*Applying Lemma 1 with $\boldsymbol{\mathcal{Y}} = 2^{[K-1]}$, we have*

$$\sum_{\boldsymbol{y} \in (2^{[K-1]})^z} \prod_{s \in \boldsymbol{y}, s \neq z} q_s \prod_{t \in [K-1] \setminus \boldsymbol{y}} (1 - q_t) = 1, \tag{4}$$

*and therefore the second term in (2) becomes*

$$\sum_{\boldsymbol{y} \in 2^{[K]}} \sum_{z \in \boldsymbol{y} \setminus \{y\}} \prod_{s \in \boldsymbol{y}, s \neq y} q_s \prod_{t \in [K] \setminus \boldsymbol{y}} (1 - q_t) w_z \psi(g_z(x)) = \sum_{z \neq y} q_z w_z \psi(g_z(x)). \tag{5}$$

*Similarly, by switching the summations, the second term on the right hand side of (1) becomes*

$$\begin{aligned}
&\beta \cdot \sum_{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}^y} \prod_{s \in \boldsymbol{y}, s \neq y} q_s \prod_{t \notin \boldsymbol{y}} (1 - q_t) \sum_{z \notin \boldsymbol{y}} w_z \psi(-g_z(x)) \\
&= \beta \cdot \sum_{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}^y} \sum_{z \notin \boldsymbol{y}} \prod_{s \in \boldsymbol{y}, s \neq y} q_s \prod_{t \notin \boldsymbol{y}} (1 - q_t) w_z \psi(-g_z(x)) \\
&= \beta \cdot \sum_{z \neq y} \sum_{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}^y \cap \boldsymbol{\mathcal{Y}}^z} \prod_{s \in \boldsymbol{y}, s \neq y} q_s \prod_{t \notin \boldsymbol{y}} (1 - q_t) w_z \psi(-g_z(x)) \\
&= \beta \cdot \sum_{z \neq y} \sum_{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}^z \setminus \boldsymbol{\mathcal{Y}}^y} \prod_{s \in \boldsymbol{y}} q_s \prod_{t \notin \boldsymbol{y}, t \neq y} (1 - q_t) w_z \psi(-g_z(x)) \\
&= \beta \cdot \sum_{z \neq y} \sum_{\boldsymbol{y} \in (2^{[K-1]})^z} \prod_{s \in \boldsymbol{y}} q_s \prod_{t \notin \boldsymbol{y}} (1 - q_t) w_z \psi(-g_z(x)) \\
&= \beta \cdot \sum_{z \neq y} q_z \sum_{\boldsymbol{y} \in (2^{[K-1]})^z} \prod_{s \in \boldsymbol{y}, s \neq z} q_s \prod_{t \notin \boldsymbol{y}} (1 - q_t) w_z \psi(-g_z(x)) \\
&= \beta \cdot \sum_{z \neq y} q_z w_z \psi(-g_z(x)),
\end{aligned} \tag{6}$$

*where the last equality holds according to (4).*

*By combining (3), (5), (6), we have*

$$\begin{aligned}
\mathcal{L}_\psi(y, g(x)) &= w_y \psi(g_y(x)) + \sum_{z \neq y} q_z w_z \psi(g_z(x)) + \beta \cdot \sum_{z \neq y} q_z w_z \psi(-g_z(x)) \\
&= w_y \psi(g_y(x)) + \sum_{z \neq y} q_z w_z \big[\psi(g_z(x)) + \beta \psi(-g_z(x))\big].
\end{aligned}$$

**Proof 5 (of Corollary 1)** *When $\beta = 0$, we naturally have $w_z = 0$ for $z \in [K] \setminus \boldsymbol{y}$, and thus there holds*

$$\begin{aligned}
\mathcal{L}_\psi(y, g(x)) &= w_y \psi(g_y(x)) + \sum_{z \neq y} w_z q_z \psi(g_z(x)) \\
&= w_y \psi(g_y(x)) + \sum_{z \in \boldsymbol{y}, z \neq y} w_z q_z \psi(g_z(x)).
\end{aligned}$$

*When $\beta = 1$, the result is obvious enough.*

*When $\beta = 2$, there holds*

$$\begin{aligned}
\mathcal{L}_\psi(y, g(x)) &= w_y \psi(g_y(x)) + \sum_{z \neq y} w_z q_z \big[\psi(g_z(x)) + 2\psi(-g_z(x))\big] \\
&= w_y \psi(g_y(x)) + \sum_{z \neq y} w_z q_z \big[\psi(g_z(x)) + \psi(-g_z(x))\big] + \sum_{z \neq y} w_z q_z \psi(-g_z(x)) \\
&= w_y \psi(g_y(x)) + \sum_{z \neq y} w_z q_z \psi(-g_z(x)) + \sum_{z \neq y} w_z q_z,
\end{aligned}$$

*where the last equation holds since $\psi(\cdot)$ is symmetric.*

# 2 Supplementary of Experiments

## 2.1 For Section 4.1

All models are trained for 500 epochs. The optimizer is stochastic gradient descent (SGD) with momentum $0.9$, with the batch size chosen as $256$. The learning rate decays by half every $50$ epochs. For our method, the initial learning rate and weight decay is chosen by five-fold cross validation from the grid $\{0.01, 0.02, 0.05, 0.1\}$ and $\{1e-3, 1e-4, 1e-5, 1e-6\}$.

## 2.2 For Section 4.3

One example of the data generation procedure in Section 4.3 is shown in the following probability matrix:

$$\begin{bmatrix}
1 & q_{adj} & q & q & \cdots & q & q_{adj} \\
q_{adj} & 1 & q_{adj} & q & \cdots & q & q \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
q_{adj} & q & q & q & \cdots & q_{adj} & 1
\end{bmatrix}$$