A The Proof of Lemma 1

Lemma 1. *Under Assumption 1, $\sum_{\bar{y}=1,\bar{y}\notin Y}^{K} \frac{\bar{p}(\boldsymbol{x},\bar{y})}{2^{K-1}-1}$ is a valid probability mass function with respect to $\boldsymbol{x}$ and Y, i.e., it is non-negative and*

$$\int_{\mathcal{X}}\int_{\mathcal{Y}} \sum_{\bar{y}=1,\bar{y}\notin Y}^{K} \frac{\bar{p}(\boldsymbol{x},\bar{y})}{2^{K-1}-1}\mathrm{d}\boldsymbol{x}\mathrm{d}Y = 1.$$

Proof. To make MLCLL hold, we remove two special subsets $Y=\emptyset$ and $Y=\mathcal{Y}$ from $\mathcal{Y}$, so let $\mathcal{Y}' = \{2^{\mathcal{Y}} - \emptyset - \mathcal{Y}\}$. Then, we have

$$\begin{aligned}
\int_{\mathcal{X}}\int_{\mathcal{Y}} \sum_{\bar{y}=1,\bar{y}\notin Y}^{K} \frac{\bar{p}(\boldsymbol{x},\bar{y})}{2^{K-1}-1}\mathrm{d}\boldsymbol{x}\mathrm{d}Y &= \int_{\mathcal{X}} \sum_{Y\in\mathcal{Y}'} \sum_{\bar{y}=1,\bar{y}\notin Y}^{K} \frac{\bar{p}(\boldsymbol{x},\bar{y})}{2^{K-1}-1}\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{X}} \sum_{\bar{y}=1}^{K} \sum_{Y\in\mathcal{Y}',\bar{y}\notin Y} \frac{\bar{p}(\boldsymbol{x},\bar{y})}{2^{K-1}-1}\mathrm{d}\boldsymbol{x} \quad \because |\{Y|Y\in\mathcal{Y}',\bar{y}\notin Y\}| = (2^{K-1}-1) \\
&= \int_{\mathcal{X}} \sum_{\bar{y}=1}^{K} \bar{p}(\boldsymbol{x},\bar{y})\mathrm{d}\boldsymbol{x} \\
&= 1.
\end{aligned}$$

□

B The Proof of Theorem 2

Theorem 2. *With $p(\boldsymbol{x},Y)$ defined in Eq. (5) and $R(\boldsymbol{f})$ defined in Eq. (1), $R(\boldsymbol{f}) = \bar{R}(\boldsymbol{f})$, where $\bar{R}(\boldsymbol{f}) = \mathbb{E}_{\bar{p}(\boldsymbol{x},\bar{y})}[\bar{\mathcal{L}}(\boldsymbol{f}(\boldsymbol{x}),\bar{y})]$ is the expected risk on complementary data, and*

$$\bar{\mathcal{L}}(\boldsymbol{f}(\boldsymbol{x}),\bar{y}) = \frac{2^{K-2}}{2^{K-1}-1}\sum_{j=1,j\neq\bar{y}}^{K} \ell_j(\boldsymbol{x}) + \frac{2^{K-2}-1}{2^{K-1}-1}\sum_{j=1,j\neq\bar{y}}^{K} \bar{\ell}_j(\boldsymbol{x}) + \bar{\ell}_{\bar{y}}(\boldsymbol{x}).$$

Proof. According to the definition of $R(\boldsymbol{f})$, we have

$$\begin{aligned}
R(\boldsymbol{f}) &= \mathbb{E}_{p(\boldsymbol{x},Y)}\left[\mathcal{L}(\boldsymbol{f}(\boldsymbol{x}),Y)\right] \\
&= \int_{\mathcal{X}} \sum_{Y\in\mathcal{Y}'} \mathcal{L}(\boldsymbol{f}(\boldsymbol{x}),Y)p(\boldsymbol{x},Y)\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{X}} \sum_{Y\in\mathcal{Y}'} \sum_{\bar{y}=1,\bar{y}\notin Y}^{K} \mathcal{L}(\boldsymbol{f}(\boldsymbol{x}),Y)\bar{p}(\boldsymbol{x},\bar{y})\mathrm{d}\boldsymbol{x} \\
&= \frac{1}{2^{K-1}-1}\int_{\mathcal{X}} \sum_{\bar{y}=1}^{K} \sum_{Y\in\mathcal{Y}',\bar{y}\notin Y} \mathcal{L}(\boldsymbol{f}(\boldsymbol{x}),Y)\bar{p}(\boldsymbol{x},\bar{y})\mathrm{d}\boldsymbol{x} \\
&= \frac{1}{2^{K-1}-1}\mathbb{E}_{\bar{p}(\boldsymbol{x},\bar{y})}\left[\sum_{Y\in\mathcal{Y}',\bar{y}\notin Y} \mathcal{L}(\boldsymbol{f}(\boldsymbol{x}),Y)\right] \\
&= \frac{1}{2^{K-1}-1}\mathbb{E}_{\bar{p}(\boldsymbol{x},\bar{y})}\left[\sum_{Y\in\mathcal{Y}',\bar{y}\notin Y} \left\{\sum_{j=1,j\in Y}^{K} \ell_j(\boldsymbol{x}) + \sum_{j=1,j\notin Y}^{K} \bar{\ell}_j(\boldsymbol{x})\right\}\right] \\
&= \frac{1}{2^{K-1}-1}\mathbb{E}_{\bar{p}(\boldsymbol{x},\bar{y})}\left[\sum_{Y\in\mathcal{Y}',\bar{y}\notin Y} \left\{\sum_{j=1,j\neq\bar{y},j\in Y}^{K} \ell_j(\boldsymbol{x}) + \sum_{j=1,j\neq\bar{y},j\notin Y}^{K} \bar{\ell}_j(\boldsymbol{x}) + \sum_{j=1,j=\bar{y},j\notin Y}^{K} \bar{\ell}_j(\boldsymbol{x})\right\}\right] \\
&= \frac{1}{2^{K-1}-1}\mathbb{E}_{\bar{p}(\boldsymbol{x},\bar{y})}\left[\sum_{j=1,j\neq\bar{y}}^{K} \sum_{Y\in\mathcal{Y}',j\in Y,\bar{y}\notin Y} \ell_j(\boldsymbol{x}) + \sum_{j=1,j\neq\bar{y}}^{K} \sum_{Y\in\mathcal{Y}',j\notin Y,\bar{y}\notin Y} \bar{\ell}_j(\boldsymbol{x}) + \sum_{Y\in\mathcal{Y}',\bar{y}\notin Y} \bar{\ell}_{\bar{y}}(\boldsymbol{x})\right]
\end{aligned}$$

$$= \frac{1}{2^{K-1}-1}\mathbb{E}_{\bar{p}(\boldsymbol{x},\bar{y})}\left[\sum_{j=1,j\neq\bar{y}}^{K} 2^{K-2}\ell_j(\boldsymbol{x}) + \sum_{j=1,j\neq\bar{y}}^{K}(2^{K-2}-1)\bar{\ell}_j(\boldsymbol{x}) + (2^{K-1}-1)\bar{\ell}_{\bar{y}}(\boldsymbol{x})\right]$$
$$= \mathbb{E}_{\bar{p}(\boldsymbol{x},\bar{y})}\left[\frac{2^{K-2}}{2^{K-1}-1}\sum_{j=1,j\neq\bar{y}}^{K}\ell_j(\boldsymbol{x}) + \frac{2^{K-2}-1}{2^{K-1}-1}\sum_{j=1,j\neq\bar{y}}^{K}\bar{\ell}_j(\boldsymbol{x}) + \bar{\ell}_{\bar{y}}(\boldsymbol{x})\right]$$
$$= \mathbb{E}_{\bar{p}(\boldsymbol{x},\bar{y})}[\bar{\mathcal{L}}(\boldsymbol{f}(\boldsymbol{x}),\bar{y})] = \bar{R}(\boldsymbol{f}).$$

□

C The Proof of Theorem 3

We start investigating an estimation error bound from the following two lemmas.

Lemma 2. *Suppose* $M = \sup_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{f}\in\mathcal{F}}\bar{\mathcal{L}}(\boldsymbol{f}(\boldsymbol{x}),\bar{y})$*,* $\mathcal{H} = \{h : (\boldsymbol{x},\bar{y}) \mapsto \bar{\mathcal{L}}(\boldsymbol{f}(\boldsymbol{x}),\bar{y}) | \boldsymbol{f} \in \mathcal{F}\}$ *is a class of measurable functions. For any* $\delta > 0$*, we are with probability at least* $1-\delta$*,*

$$\sup_{\boldsymbol{f}\in\mathcal{F}}|\bar{R}_n(\boldsymbol{f}) - \bar{R}(\boldsymbol{f})| \leq 2\mathfrak{R}_n(\mathcal{H}) + \frac{M}{2}\sqrt{\frac{\log 2/\delta}{2n}},$$

where $\mathfrak{R}_n(\mathcal{H}) = \mathbb{E}_{\boldsymbol{x},\bar{y},\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i h(\boldsymbol{x}_i,\bar{y}_i)\right]$ *is the expected Rademacher complexity of* $\mathcal{H}$*, in which* $\boldsymbol{\sigma} = \{\sigma_1,\ldots,\sigma_n\}$ *are* n *Rademacher variables.*

Proof. To proof this lemma, we show that the single direction $\sup_{\boldsymbol{f}\in\mathcal{F}}(\bar{R}_n(\boldsymbol{f}) - \bar{R}(\boldsymbol{f}))$ is bounded with the probability $1-\sigma/2$. Based on $\bar{\mathcal{L}}(\boldsymbol{f}(\boldsymbol{x}),\bar{y})$, suppose an instance $(\boldsymbol{x}_i,\bar{y}_i)$ is replaced by an arbitrary instance $(\boldsymbol{x}'_i,\bar{y}'_i)$, whose change is no greater than $M/2n$. By applying McDiarmid's inequality, for any $\delta > 0$, with the probability at least $1-\delta/2$, we have

$$\sup_{\boldsymbol{f}\in\mathcal{F}}(\bar{R}_n(\boldsymbol{f}) - \bar{R}(\boldsymbol{f})) \leq \mathbb{E}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}(\bar{R}_n(\boldsymbol{f}) - \bar{R}(\boldsymbol{f}))\right] + \frac{M}{2}\sqrt{\frac{\log 2/\delta}{2n}}.$$

By sysmmetrization, we have

$$\mathbb{E}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\left(\bar{R}_n(\boldsymbol{f}) - \bar{R}(\boldsymbol{f})\right)\right] \leq 2\mathfrak{R}_n(\mathcal{H}).$$

□

Then, we give an upper bound of $\mathfrak{R}_n(\mathcal{H})$.

Lemma 3. *For any* $j \in \mathcal{Y}$*, assuming* $\ell_j(\boldsymbol{x})$ *and* $\bar{\ell}_j(\boldsymbol{x})$ *are* β^+*-Lipschitz and* β^-*-Lipschitz with respect to* $\boldsymbol{f}(\boldsymbol{x})$ *respectively, then we have*

$$\mathfrak{R}_n(\mathcal{H}) \leq \sqrt{2}\left[\frac{(K-1)2^{K-2}}{2^{K-1}-1}\beta^+ + \frac{(K-1)2^{K-2}-K}{2^{K-1}-1}\beta^-\right]\sum_{j=1}^{K}\mathfrak{R}_n(\mathcal{G}_j),$$

where $\mathcal{G}_j = \{g : \boldsymbol{x} \mapsto f_j(\boldsymbol{x}) | \boldsymbol{f} \in \mathcal{F}\}$ *and* $\mathfrak{R}_n(\mathcal{G}_j) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{\sigma}}\left[\sup_{g\in\mathcal{G}_j}\frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{x}_i)\right]$.

Proof. Firstly, suppose $\ell\circ\mathcal{F}$ and $\bar{\ell}\circ\mathcal{F}$ denote $\{\ell\circ\mathcal{F}|\boldsymbol{f}\in\mathcal{F}\}$ and $\{\bar{\ell}\circ\mathcal{F}|\boldsymbol{f}\in\mathcal{F}\}$ respectively. Then we apply the Rademacher vector contraction inequality [Maurer, 2016]:

$$\mathfrak{R}_n(\mathcal{H}) = \mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{h\in\mathcal{H}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i h(\boldsymbol{x}_i,\bar{y}_i)\right]$$
$$= \mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\bar{\mathcal{L}}(\boldsymbol{f}(\boldsymbol{x}_i),\bar{y}_i)\right]$$
$$= \mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\left\{\frac{2^{K-2}}{2^{K-1}-1}\sum_{j=1,j\neq\bar{y}_i}^{K}\ell_j(\boldsymbol{x}_i) + \frac{2^{K-2}-1}{2^{K-1}-1}\sum_{j=1,j\neq\bar{y}_i}^{K}\bar{\ell}_j(\boldsymbol{x}_i) + \bar{\ell}_{\bar{y}_i}(\boldsymbol{x}_i)\right\}\right]$$

$$
\begin{aligned}
&\leq \frac{2^{K-2}}{2^{K-1}-1}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\sum_{j=1,j\neq\bar{y}_i}^{K}\ell_j(\boldsymbol{x}_i)\right] + \frac{2^{K-2}-1}{2^{K-1}-1}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\sum_{j=1,j\neq\bar{y}_i}^{K}\bar{\ell}_j(\boldsymbol{x}_i)\right] + \\
&\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\bar{\ell}_{\bar{y}_i}(\boldsymbol{x}_i)\right] \\
&\leq \frac{2^{K-2}}{2^{K-1}-1}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\sum_{j=1}^{K}\ell_j(\boldsymbol{x}_i)\right] - \frac{2^{K-2}}{2^{K-1}-1}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\ell_{\bar{y}_i}(\boldsymbol{x}_i)\right] \\
&+ \frac{2^{K-2}-1}{2^{K-1}-1}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\sum_{j=1}^{K}\bar{\ell}_j(\boldsymbol{x}_i)\right] - \frac{2^{K-2}-1}{2^{K-1}-1}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\bar{\ell}_{\bar{y}_i}(\boldsymbol{x}_i)\right] + \mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\bar{\ell}_{\bar{y}_i}(\boldsymbol{x}_i)\right] \\
&\leq \frac{2^{K-2}}{2^{K-1}-1}\sum_{j=1}^{K}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\ell_j(\boldsymbol{x}_i)\right] - \frac{2^{K-2}}{2^{K-1}-1}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\ell_{\bar{y}_i}(\boldsymbol{x}_i)\right] \\
&+ \frac{2^{K-2}-1}{2^{K-1}-1}\sum_{j=1}^{K}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\bar{\ell}_j(\boldsymbol{x}_i)\right] + \frac{2^{K-2}}{2^{K-1}-1}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{f}\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sigma_i\bar{\ell}_{\bar{y}_i}(\boldsymbol{x}_i)\right] \\
&\leq \frac{2^{K-2}}{2^{K-1}-1}\sum_{j=1}^{K}\mathfrak{R}_n(\ell\circ\mathcal{F}) - \frac{2^{K-2}}{2^{K-1}-1}\mathfrak{R}_n(\ell\circ\mathcal{F}) + \frac{2^{K-2}-1}{2^{K-1}-1}\sum_{j=1}^{K}\mathfrak{R}_n(\bar{\ell}\circ\mathcal{F}) + \frac{2^{K-2}}{2^{K-1}-1}\mathfrak{R}_n(\bar{\ell}\circ\mathcal{F}) \\
&= \frac{2^{K-2}(K-1)}{2^{K-1}-1}\mathfrak{R}_n(\ell\circ\mathcal{F}) + \frac{2^{K-2}(K-1)-K}{2^{K-1}-1}\mathfrak{R}_n(\bar{\ell}\circ\mathcal{F}) \\
&\leq \frac{2^{K-2}(K-1)}{2^{K-1}-1}\sqrt{2}\beta^{+}\sum_{j=1}^{K}\mathfrak{R}_n(\mathcal{G}_j) + \frac{2^{K-2}(K-1)-K}{2^{K-1}-1}\sqrt{2}\beta^{-}\sum_{j=1}^{K}\mathfrak{R}_n(\mathcal{G}_j) \\
&= \sqrt{2}\left[\frac{2^{K-2}(K-1)}{2^{K-1}-1}\beta^{+} + \frac{2^{K-2}(K-1)-K}{2^{K-1}-1}\beta^{-}\right]\sum_{j=1}^{K}\mathfrak{R}_n(\mathcal{G}_j).
\end{aligned}
$$

□

Based on Lemma 2 and 3, we can derive the estimation error bound.

Theorem 3. *Suppose $M = \sup_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{f}\in\mathcal{F}}\bar{\mathcal{L}}(\boldsymbol{f}(\boldsymbol{x}),\bar{y})$. For any $j\in\mathcal{Y}$, assuming $\ell_j(\boldsymbol{x})$ and $\bar{\ell}_j(\boldsymbol{x})$ are β^{+}-Lipschitz and β^{-}-Lipschitz with respect to $\boldsymbol{f}(\boldsymbol{x})$ respectively. For any δ, with the probability at least $1-\delta$,*

$$
R(\boldsymbol{f}_n) - R(\boldsymbol{f}^*) \leq M\sqrt{\frac{\log 2/\delta}{2n}} + 4\sqrt{2}\left[\frac{(K-1)2^{K-2}}{2^{K-1}-1}\beta^{+} + \frac{(K-1)2^{K-2}-K}{2^{K-1}-1}\beta^{-}\right]\sum_{j=1}^{K}\mathfrak{R}_n(\mathcal{G}_j).
$$

Proof. Combining Theorem 3, Lemma 2 and 3, the following inequality holds:

$$
\begin{aligned}
R(\boldsymbol{f}_n) - R(\boldsymbol{f}^*) &= \bar{R}(\boldsymbol{f}_n) - \bar{R}(\boldsymbol{f}^*) \\
&= [\bar{R}(\boldsymbol{f}_n) - \bar{R}_n(\boldsymbol{f}^*)] + [\bar{R}_n(\boldsymbol{f}_n) - \bar{R}_n(\boldsymbol{f}^*)] + [\bar{R}_n(\boldsymbol{f}^*) - \bar{R}(\boldsymbol{f}^*)] \\
&\leq \bar{R}(\boldsymbol{f}_n) - \bar{R}_n(\boldsymbol{f}_n) + \bar{R}_n(\boldsymbol{f}^*) - \bar{R}(\boldsymbol{f}^*) \\
&= 2\sup_{\boldsymbol{f}\in\mathcal{F}}|\bar{R}_n(\boldsymbol{f}) - \bar{R}(\boldsymbol{f})| \\
&\leq M\sqrt{\frac{\log 2/\delta}{2n}} + 4\sqrt{2}\left[\frac{(K-1)2^{K-2}}{2^{K-1}-1}\beta^{+} + \frac{(K-1)2^{K-2}-K}{2^{K-1}-1}\beta^{-}\right]\sum_{j=1}^{K}\mathfrak{R}_n(\mathcal{G}_j).
\end{aligned}
$$

□

Table 3: Characteristics of datasets. $dim(\mathcal{S})$, $|\mathcal{S}|$ and $L(\mathcal{S})$ refer to the number of features, instances and labels respectively. $LCard(\mathcal{S})$ and avg. #CL denote the average number of labels and complemenatry labels per instance respectively.

Datasets	$dim(\mathcal{S})$	$\|\mathcal{S}\|$	$L(\mathcal{S})$	$LCard(\mathcal{S})$	avg. #CL
scene	294	2407	6	1.07	1
yeast	103	2417	14	4.23	1
bookmark	2150	38912	15	1.25	1
mediamill	120	41701	15	3.63	1
eurlex_dc	100	8636	15	1.02	1
		19340	410	1.29	205
eurlex_sm	100	13270	15	1.74	1
		19338	201	2.21	100
Corel16k	500	11103	15	1.77	1
		13766	153	2.87	77
delicious	500	14784	15	4.32	1
		16105	983	19.02	492

D Details of Experiments

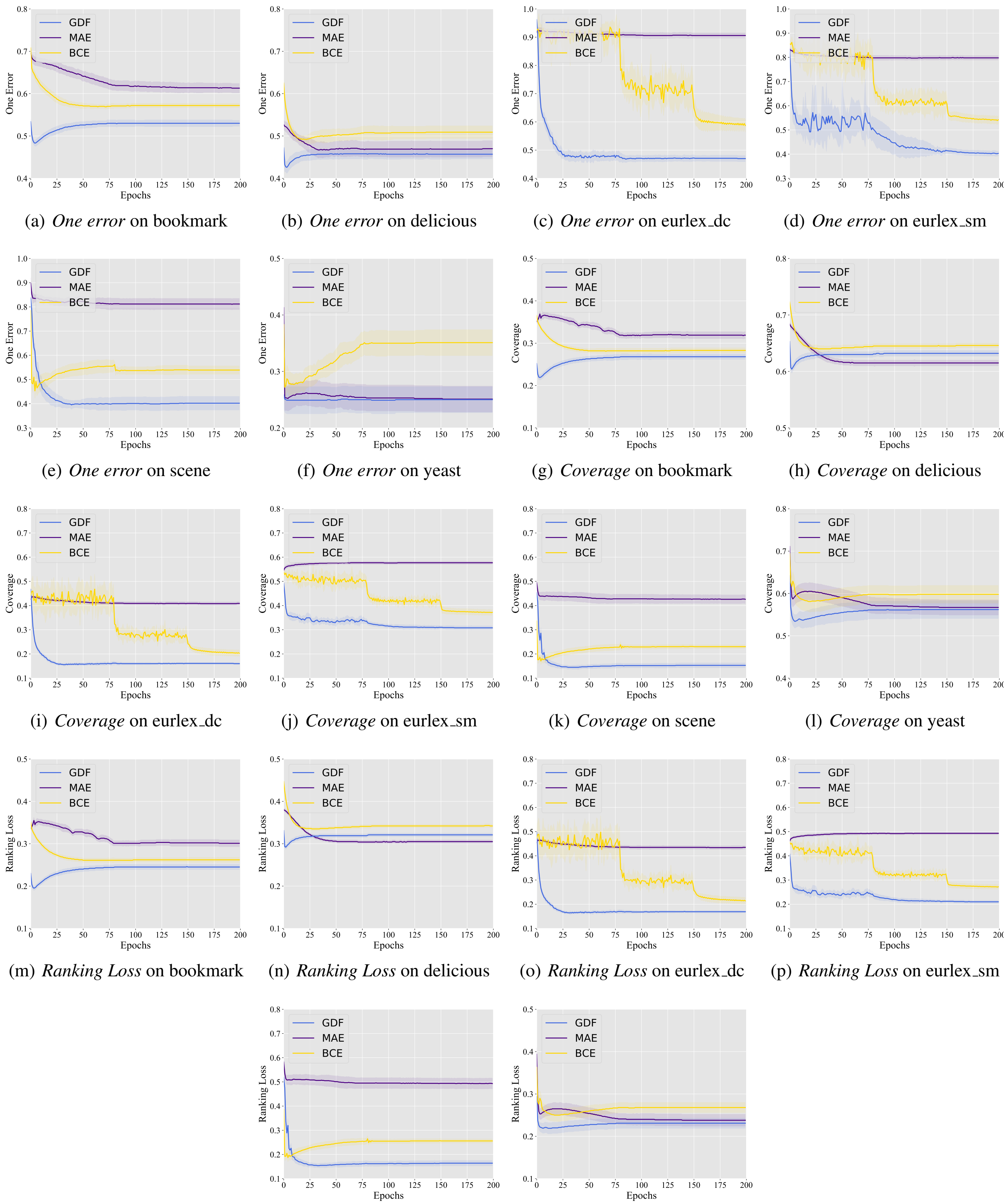
Details of datasets are presented in Table 3, including the number of features ($dim(\mathcal{S})$), the number of instances ($|\mathcal{S}|$), the number of labels ($L(\mathcal{S})$), the average number of relevant labels per instance ($LCard(\mathcal{S})$), and the average number of complementary labels per instance (avg. #CL). Besides, Table 4 and 5 are shown results of various approaches on different datasets that are applied the first pre-processing way and second pre-processing way to process respectively. As can be seen from Table 4 and 5, our approach can perform well in most cases. Accordingly, we draw the curves of our empirical risk estimator with MAE loss and BCE loss, and GDF loss on *one error*, *coverage*, and *ranking loss*. Due to the results of *hamming loss* relying on the threshold value, its curve does not be presented here. From Fig. 3, GDF achieves a good performance in most cases, which demonstrates the superiority of our proposed approach.

Table 4: *Hamming loss* and *coverage* (mean±std) on training data with first pre-processing way (each instance is given one complementary label and labels are kept under 15). The best performance of each dataset is presented in **boldface**, where •/○ denotes whether GDF is superior/inferior to baselines with pairwise t-test (at 0.05 significance level).

Methods	MLL		PML		CLL	Unbounded	Bounded	GDF
	ML-KNN	LIFT	fpml	PML-LRS	L-UW	BCE	MAE	
Hamming Loss↓								
bookmark	.917±.001•	.916±.001•	.420±.009•	.813±.003•	.295±.007•	**.147±.002**	.237±.006•	.149±.004
Corel16k	.882±.008•	.882±.008•	.882±.008•	.862±.001•	.365±.021•	.240±.006•	.200±.013	**.196±.019**
delicious	.711±.003•	.711±.003•	.712±.003•	.459±.002•	.331±.005•	.300±.002•	.285±.004	**.283±.003**
mediamill	.758±.011•	.758±.011•	.757±.011•	.760±.000•	.185±.016	**.173±.011**	.179±.016	.180±.011
eurlex_dc	.932±.000•	.932±.000•	.178±.012○	.777±.008•	.541±.014•	**.136±.004**○	.513±.013•	.225±.009
eurlex_sm	.883±.001•	.883±.001•	.178±.005○	.711±.002•	.533±.014•	**.176±.012**○	.500±.006•	.211±.004
scene	.821±.002•	.820±.003•	.819±.002•	.814±.000•	.526±.016•	.203±.014	.372±.009•	**.202±.010**
yeast	.697±.012•	.697±.012•	.697±.012•	.316±.000•	.287±.010•	.247±.008•	.240±.008•	**.231±.007**
Coverage↓								
bookmark	.359±.006•	.328±.008•	.474±.018•	.280±.004•	.219±.007	.281±.006•	.318±.008•	**.219±.006**
Corel16k	.430±.042	.487±.026•	.513±.034•	**.404±.008**	.405±.037	.493±.020•	.420±.031	.410±.039
delicious	.712±.006•	.703±.004•	.726±.009•	.609±.003	.634±.006•	.641±.005•	.615±.006•	**.605±.007**
mediamill	.503±.022•	.501±.023•	.512±.034•	**.436±.002**○	.483±.033	.449±.022	.494±.034•	.463±.023
eurlex_dc	.266±.007•	.277±.011•	.441±.035•	.170±.005•	.252±.014•	.205±.016•	.411±.008•	**.153±.005**
eurlex_sm	.411±.008•	.421±.012•	.552±.031•	.336±.005•	.423±.008•	.374±.017•	.546±.007•	**.308±.008**
scene	.299±.025•	.256±.016•	.434±.02•	.230±.006•	.397±.015•	.169±.017•	.425±.019•	**.144±.008**
yeast	.579±.017•	.649±.019•	.553±.031	.742±.025•	.585±.017•	.582±.024•	.567±.018•	**.540±.020**

Table 5: Experimental results (mean±std) on the training data adopted the second pre-processing way. The label space is original and the number of complementary labels per instance is half the original number of labels. The best performance of each dataset is shown in **boldface**, where •/○ indicates whether GDF is superior/inferior to baselines with pairwise t-test (at 0.05 significance level).

Methods	MLL	PML		CLL	Bounded	GDF
	ML-KNN	fpml	PML-LRS	L-UW	MAE	
Hamming Loss↓						
Corel16k	.299±.004•	.195±.011•	.452±.003•	.036±.002•	.060±.003•	.029±.003
delicious	.391±.003•	.262±.009•	.202±.001•	.025±.000•	.038±.001•	.019±.000
eurlex_dc	.474±.007•	.091±.005•	.136±.003•	.271±.015•	.196±.006•	.020±.001
eurlex_sm	.483±.008•	.066±.005•	.146±.002•	.312±.019•	.233±.006•	.019±.001
One Error↓						
Corel16k	.789±.050•	.770±.052	.721±.014	.769±.053	.742±.048	.736±.053
delicious	.541±.018•	.593±.015•	.455±.012•	.526±.006•	.510±.009•	.383±.014
eurlex_dc	.716±.011○	.898±.011•	.722±.002	.804±.007•	.815±.009•	.728±.009
eurlex_sm	.535±.011•	.778±.007•	.525±.003	.599±.009•	.660±.009•	.523±.007
Coverage↓						
Corel16k	.507±.070	.458±.091	.519±.019	.451±.090	.449±.088	.519±.061
delicious	.845±.004•	.803±.008•	.803±.011•	.753±.005○	.746±.005○	.767±.003
eurlex_dc	.310±.005•	.305±.006•	.296±.004•	.293±.006•	.301±.010•	.283±.006
eurlex_sm	.285±.007•	.325±.007•	.278±.004•	.287±.006•	.273±.006•	.268±.008



(a) *One error* on bookmark (b) *One error* on delicious (c) *One error* on eurlex_dc (d) *One error* on eurlex_sm

(e) *One error* on scene (f) *One error* on yeast (g) *Coverage* on bookmark (h) *Coverage* on delicious

(i) *Coverage* on eurlex_dc (j) *Coverage* on eurlex_sm (k) *Coverage* on scene (l) *Coverage* on yeast

(m) *Ranking Loss* on bookmark (n) *Ranking Loss* on delicious (o) *Ranking Loss* on eurlex_dc (p) *Ranking Loss* on eurlex_sm

(q) *Ranking Loss* on scene (r) *Ranking Loss* on yeast

Figure 3: *One error*, *coverage*, and *ranking loss* on various datasets adopted the first pre-processing way, where each instance is associated with one complementary label and labels are kept under 15. Dark colors show the mean of testing average precision and light colors are corresponding to the std.