Contents

A. More Discussions on EM

A.1. EM Basics

The Expectation–Maximization (EM) algorithm is an iterative method used to find MLE or MAP estimates $\boldsymbol{\pi}^*$ of parameters $\boldsymbol{\pi}$. That is

$$\boldsymbol{\pi}^* = \arg\max_{\boldsymbol{\pi}} L(\boldsymbol{\pi}; X),$$

where $L(\boldsymbol{\pi}; X)$ is a log posterior of $\boldsymbol{\pi}$ given X or a log-likelihood of X given $\boldsymbol{\pi}$. The algorithm works when the statistical model has unobserved latent variables Y.

An EM algorithm consists of two steps in every iteration, namely the **Expectation-Step** and **Maximization-Step**, which are usually referred as the **E-Step** and **M-Step** respectively. To derive an EM algorithm that maximizes $L(\boldsymbol{\pi}; X)$, we first write the analytical expression of the log posterior or log likelihood of parameter $\boldsymbol{\pi}$ given observed variable X and unobserved latent variable Z as $L(\boldsymbol{\pi}; X, Y)$.

In the **E-Step**, the model constructs a $Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})$ as the expectation of $L(\boldsymbol{\pi}; X, Y)$ w.r.t latent variable Y given observed variable X and current $\boldsymbol{\pi}^{(t)}$:

$$Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)}) = \mathop{\mathbb{E}}_{Y|X,\boldsymbol{\pi}^{(t)}}[L(\boldsymbol{\pi}; X, Y)]. \tag{21}$$

In the **M-Step**, find the optimal $\boldsymbol{\pi}^{(t+1)}$ with:

$$\boldsymbol{\pi}^{(t+1)} = \arg\max_{\boldsymbol{\pi}} Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)}). \tag{22}$$

By repeating the two steps until convergence, under mild conditions, the algorithm is guaranteed to converge to a stationary point of $L(\boldsymbol{\pi}; X)$. We recommend this lecture note for more detailed proof and discussion.

A.2. EM Advantages

Some advantages of the EM algorithm are:

- The latent variable may be difficult to integrate out. The EM-algorithm allows one to manipulate conditional distributions by marginalising out a certain log likelihood with respect to the probability measure of a latent variable conditioned on data and parameters, which is often easier than integrating the latent variable directly.
- The **M-Step** often admits a closed form solution.
- It is guaranteed that $L(\boldsymbol{\pi}; X) - L(\boldsymbol{\pi}^{(t)}; X) \geq Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)}) - Q(\boldsymbol{\pi}^{(t)}|\boldsymbol{\pi}^{(t)})$, which means improvement in $L(\boldsymbol{\pi}; X)$ won't be less than improvement in $Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})$ by solve for best $\boldsymbol{\pi}^{(t+1)}$ in the t^{th} **M-Step**.

B. Mathematical Proofs

B.1. Proof of Convexity of MAP objective

Proof. We assume the parameter $\boldsymbol{\pi}$ of target label distribution $\mathbb{P}(Y_t = i|\boldsymbol{\pi}) = \pi_j$ follows a prior distribution $\boldsymbol{\pi} \sim \text{Dir}(K, \boldsymbol{\alpha})$, which is a Dirichlet distribution. $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ...\alpha_K]$ is the parameter of the prior. For unlabelled sample $\mathbb{X} = \{x_i | i = 1, 2, ...N\}$ drawn i.i.d from target distribution P_t, we want to find parameter $\boldsymbol{\pi} \in \Delta^{K-1}$ that maximizes the posterior $\mathbb{P}(\boldsymbol{\pi}|\mathbb{X}, \boldsymbol{\alpha})$, or equivalently minimizes the negative log posterior:

$$\boldsymbol{\pi}^* = \underset{\boldsymbol{\pi}\in\Delta^{K-1}}{\arg\min} - \log \mathbb{P}(\boldsymbol{\pi}|\mathbb{X}, \boldsymbol{\alpha}) = \underset{\boldsymbol{\pi}\in\Delta^{K-1}}{\arg\min} - \left(\log \prod_{i=1}^{N} \mathbb{P}(X_t = x_i|\boldsymbol{\pi}) + \log \mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha}) + Const \right) \tag{23}$$

where $Const$ includes all terms that can be treated as a constant w.r.t $\boldsymbol{\pi}$. $\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})$ is the Dirichlet prior, with $\log \mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})$ as a concave function of $\boldsymbol{\pi}$:

$$\log \mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha}) = \log \frac{1}{\mathrm{B}(\boldsymbol{\alpha})} \prod_{i=1}^{K} \pi_i^{\alpha_i - 1} = \sum_{i=1}^{K} (\alpha_i - 1) \log \pi_i + Const \tag{24}$$

with $\mathrm{B}(\boldsymbol{\alpha})$ as the normalization constant for a given $\boldsymbol{\alpha}$. Here we require $\alpha_i - 1 > 0, i = 1, 2...K$.

Suppose the equality in (25) below holds for the source label distribution $\mathbb{P}(Y_s = j)$, classifier f and target label distribution $\mathbb{P}(Y_t = j|\boldsymbol{\pi})$:

$$\begin{aligned} \mathbb{P}(Y_s = j) &= c_j > 0 \\ \mathbb{P}(Y_s = j|X_s = x) &= f(x)_j \\ \mathbb{P}(Y_t = j|\boldsymbol{\pi}) &= \pi_j \end{aligned} \tag{25}$$

Alexandari *et al.* [1] has proved that under label shift, if (25) holds for $(x, i) \in \mathcal{X} \times \mathcal{Y}$, then the negative log likelihood defined in (26) is convex.

$$-\log L(\boldsymbol{\pi}; \mathbb{X}) = -\log \mathbb{P}(\mathbb{X}|\boldsymbol{\pi}) = -\log \prod_{i=1}^{N} \mathbb{P}(X_t = x_i|\boldsymbol{\pi}) \tag{26}$$

Thus the MAP objective in (23) can be rewritten as:

$$\boldsymbol{\pi}^* = \underset{\boldsymbol{\pi}\in\Delta^{K-1}}{\arg\min} - \left(\log L(\boldsymbol{\pi}; \mathbb{X}) + \sum_{i=1}^{K} (\alpha_i - 1) \log \pi_i + Const \right). \tag{27}$$

The objective function in (27) is adding (26) with extra term $\sum_{l=1}^{K} (\alpha_l - 1) \log \pi_l$, which is a strictly convex function. Given the constraints $\boldsymbol{\pi} \in \Delta^{K-1} \subseteq \mathbb{R}^K$ defines a convex set, the above optimization problem is a convex optimization problem with a unique global minima. ■

B.2. Proof of EM for MAP estimation

Proof. We follow the standard EM derivation procedure. We first derive the form of complete posterior $\mathbb{P}(\boldsymbol{\pi}|\mathbb{X},\mathbb{Y},\boldsymbol{\alpha})$ with unobserved latent variable $\mathbb{Y}$ based on the original posterior $\mathbb{P}(\boldsymbol{\pi}|\mathbb{X},\boldsymbol{\alpha})$. Based on same assumptions made by MLLS [1,32], we construct analytical expression of $Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})$ function for E-Step in EM. Finally, we optimize $Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})$ w.r.t $\boldsymbol{\pi}$ to find the M-Step.

The EM algorithm optimizes the original posterior $\mathbb{P}(\boldsymbol{\pi}|\mathbb{X},\boldsymbol{\alpha})$ with an iterative procedure. Target labels $\mathbb{Y} = \{y_i|\{y_i,x_i\} \sim P_t, x_i \in \mathbb{X}\}$ of input image $\mathbb{X}$ are treated as unobserved latent variables. The complete posterior $\mathbb{P}(\boldsymbol{\pi}|\mathbb{X},\mathbb{Y},\boldsymbol{\alpha})$ that includes latent variable $\mathbb{Y}$ can be written as:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{\pi}|\mathbb{X},\mathbb{Y},\boldsymbol{\alpha}) &= \frac{\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})\mathbb{P}(\mathbb{X},\mathbb{Y}|\boldsymbol{\pi})}{\int_{\boldsymbol{\pi}}\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})\mathbb{P}(\mathbb{X},\mathbb{Y}|\boldsymbol{\pi})d\boldsymbol{\pi}} \\
&= \frac{1}{Z}\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})\prod_{i=1}^{N}\mathbb{P}(X_t=x_i,Y_t=y_i|\boldsymbol{\pi}) \\
&= \frac{1}{Z}\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})\prod_{i=1}^{N}\mathbb{P}(X_t=x_i|Y_t=y_i)\mathbb{P}(Y_t=y_i|\boldsymbol{\pi}) \\
&= \frac{1}{Z}\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})\prod_{i=1}^{N}\mathbb{P}(X_s=x_i|Y_s=y_i)\mathbb{P}(Y_t=y_i|\boldsymbol{\pi}) \\
&= \frac{1}{Z}\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})\prod_{i=1}^{N}\frac{\mathbb{P}(Y_t=y_i|\boldsymbol{\pi})}{\mathbb{P}(Y_s=y_i)}\mathbb{P}(Y_s=y_i|X_s=x_i)\mathbb{P}(X_s=x_i)
\end{aligned} \tag{28}$$

where Z is the integral in the denominator that has integrated out $\boldsymbol{\pi}$ and can be treated as a constant w.r.t $\boldsymbol{\pi}$. Further, $\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})$ is the Dirichlet prior of $\boldsymbol{\pi}$ that has the expression:

$$\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha}) = \frac{1}{\mathrm{B}(\boldsymbol{\alpha})}\prod_{i=1}^{K}\pi_i^{\alpha_i-1} \tag{29}$$

with $\mathrm{B}(\boldsymbol{\alpha})$ as the normalization constant for a given $\boldsymbol{\alpha}$ and $\alpha_i - 1 > 0, i = 1,2...K$.

In the **E-Step**, given $\boldsymbol{\pi}^{(t)}$ in the t^{th} iteration, with the complete posterior $\mathbb{P}(\boldsymbol{\pi}|\mathbb{X},\mathbb{Y},\boldsymbol{\alpha})$, the $Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})$ can be written as:

$$\begin{aligned}
Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)}) &= \mathop{\mathbb{E}}_{\mathbb{Y}|\mathbb{X},\boldsymbol{\pi}^{(t)}}\left[\log\mathbb{P}(\boldsymbol{\pi}|\mathbb{X},\mathbb{Y},\boldsymbol{\alpha})\right] \\
&= \mathop{\mathbb{E}}_{\mathbb{Y}|\mathbb{X},\boldsymbol{\pi}^{(t)}}\left[\log\left(\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})\prod_{i=1}^{N}\frac{\mathbb{P}(Y_t=y_i|\boldsymbol{\pi})}{\mathbb{P}(Y_s=y_i)}\mathbb{P}(Y_s=y_i|X_s=x_i)\mathbb{P}(X_s=x_i)\right)+Const\right] \\
&= \mathop{\mathbb{E}}_{\mathbb{Y}|\mathbb{X},\boldsymbol{\pi}^{(t)}}\left[\sum_{i=1}^{N}\log\prod_{j=1}^{K}\pi_j^{\mathbb{I}(y_i=j)}\right]+\sum_{l=1}^{K}(\alpha_l-1)\log\pi_l+Const \\
&= \sum_{i=1}^{N}\sum_{j=1}^{K}\mathop{\mathbb{E}}_{\mathbb{Y}|\mathbb{X},\boldsymbol{\pi}^{(t)}}\left[\mathbb{I}(y_i=j)\right]\log\pi_j+\sum_{l=1}^{K}(\alpha_l-1)\log\pi_l+Const \\
&= \sum_{j=1}^{K}\sum_{i=1}^{N}\mathbb{P}(Y_t=j|X_t=x_i,\boldsymbol{\pi}^{(t)})\log\pi_j+\sum_{l=1}^{K}(\alpha_l-1)\log\pi_l+Const
\end{aligned} \tag{30}$$

where $\mathbb{I}(y_i = j)$ is the indicator function. In the above derivation, all terms that are irrelevant to $\boldsymbol{\pi}$ are moved in the term $Const$. For example, $\mathbb{E}_{\mathbb{Y}|\mathbb{X},\boldsymbol{\pi}^{(t)}}[\mathbb{P}(Y_s=y_i)]$ and $\log\mathrm{B}(\boldsymbol{\alpha})$.

Note that in label shift, Saren *et al.* [32] proved that under label shift (1), $\mathbb{P}(Y_t=j|X_t=x_i,\boldsymbol{\pi})$ can be written as:

$$\mathbb{P}(Y_t=y_i|X_t=x_i,\boldsymbol{\pi}) = \frac{\frac{\mathbb{P}(Y_t=y_i|\boldsymbol{\pi})}{\mathbb{P}(Y_s=y_i)}\mathbb{P}(Y_s=y_i|X_s=x_i)}{\sum_{l=1}^{K}\frac{\mathbb{P}(Y_t=l|\boldsymbol{\pi})}{\mathbb{P}(Y_s=l)}\mathbb{P}(Y_s=l|X_s=x_i)} \tag{31}$$

We can substitute analytical expression of each probability into the equation with (25):

$$g(x_i, \boldsymbol{\pi})_j := \mathbb{P}(Y_t = j | X_t = x_i, \boldsymbol{\pi}) = \frac{\frac{\pi_j}{c_j} f(x_i)_j}{\sum_{l=1}^{K} \frac{\pi_l}{c_l} f(x_i)_l} \tag{32}$$

Then the $Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})$ can be rewritten as:

$$Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)}) = \sum_{j=1}^{K} \sum_{i=1}^{N} g(x_i; \boldsymbol{\pi}^{(t)})_j \log \pi_j + \sum_{l=1}^{K} (\alpha_l - 1) \log \pi_l + Const \tag{33}$$

In the **M-step**, we solve the optimization objective with respect to $\boldsymbol{\pi}$:

$$\boldsymbol{\pi}^{(t+1)} = \underset{\boldsymbol{\pi} \in \Delta^{K-1}}{\arg\max} Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)}) \tag{34}$$

By substitution, the objective can be rewritten as:

$$\begin{cases} \min\limits_{\boldsymbol{\pi}} - \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{K} g(x_i; \boldsymbol{\pi}^{(t)})_j \log \pi_j - \sum\limits_{l=1}^{K} (\alpha_l - 1) \log \pi_l \\ \text{s.t: } \sum\limits_{j=1}^{K} \pi_j = 1 \text{ and } \pi_i \geq 0, i \in [1, 2, ...K] \end{cases} \tag{35}$$

Convexity The objective we want to optimize is just a linear combination of $\log \pi_i$, which is a concave function w.r.t $\boldsymbol{\pi}$. Knowing that the constraints define a convex set on $\mathbb{R}^K$, the above optimization problem is also a convex optimization problem and every local minima is a global minima.

Optimization without inequality constraints With only equality constraints, standard Lagrangian Multiplier method can be applied. The Lagrangian can be written as:

$$\mathcal{L}(\boldsymbol{\pi}, \lambda) = \sum_{i=1}^{N} \sum_{j=1}^{K} g(x_i; \boldsymbol{\pi}^{(t)})_j \log \pi_j + \sum_{j=1}^{K} (\alpha_j - 1) \log \pi_j + \lambda (1 - \sum_{j=1}^{K} \pi_j) \tag{36}$$

The optimal $\boldsymbol{\pi}$ can be found by taking all the partial derivative of $\mathcal{L}(\boldsymbol{\pi}, \lambda)$ w.r.t π_j and λ to 0:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \pi_j} = \dfrac{\sum_{i=1}^{N} g(x_i; \boldsymbol{\pi}^{(t)})_j}{\pi_j} + \dfrac{\alpha_j - 1}{\pi_j} - \lambda = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda} = \sum\limits_{i=1}^{K} \pi_i - 1 = 0 \end{cases} \tag{37}$$

The solution to the above equation set can be written as:

$$\begin{cases} \pi_j = \dfrac{\sum_{i=1}^{N} g(x_i; \boldsymbol{\pi}^{(t)})_j + \alpha_j - 1}{\lambda} \\ \lambda = N + \sum\limits_{l=1}^{K} (\alpha_l - 1) \end{cases} \tag{38}$$

Therefore optimal $\boldsymbol{\pi}$ for $Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})$ without inequality constraints is given by:

$$\pi_j = \frac{\sum_{i=1}^{N} g(x_i; \boldsymbol{\pi}^{(t)})_j + \alpha_j - 1}{N + \sum_{l=1}^{K} (\alpha_l - 1)} \tag{39}$$

Proof that the solution satisfies inequality constraints In the expression of $g(x_i;\boldsymbol{\pi}^{(t)})_j$ defined in 32, the output of the classifier $f:\mathcal{X}\to\Delta^{K-1}$ is a probability simplex and thus is non-negative. Note that we also have $c_i>0, i=1,2...K$. Therefore $\boldsymbol{\pi}^{(t)}>0 \Rightarrow g(x_i;\boldsymbol{\pi}^{(t)})_j>0$ is non-negative. Note that we also require $\alpha_i-1>0, i=1,2...K$ when defining the Dirichlet prior. Therefore we have $\boldsymbol{\pi}^{(t)}>0\Rightarrow\boldsymbol{\pi}^{(t+1)}>0$.

Because the optimization problem is convex, when $\pi_j^{(t)}>0, j=1,2,...K$, the above equation gives the global optimal $\boldsymbol{\pi}^{(t+1)}$:

$$\pi_j^{(t+1)}=\left(\arg\max_{\boldsymbol{\pi}\in\Delta^{K-1}} Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})\right)_j=\frac{\sum_{i=1}^N g(x_i;\boldsymbol{\pi}^{(t)})_j+\alpha_j-1}{N+\sum_{l=1}^K(\alpha_l-1)} \tag{40}$$

■

Algorithm 3 MAPLS (Formal EM Formulation)

Input: Target domain unlabeled data $\{x_i|i=1,2,...N,\{x_i,\cdot\}\sim P_t\}$, source domain label distribution $P(Y_s=j)=c_j$ and blackbox classifier $f:\mathcal{X}\to\Delta^{K-1}$, Dirichlet prior $\mathbb{P}(\boldsymbol{\pi}|\boldsymbol{\alpha})$.

Initialize: $\boldsymbol{\pi}^{(0)}\in\Delta^{K-1}$ with $\pi_i^{(0)}>0, i=1,2...K$

for $t=0$ to T **do**

Estimating latent conditional distribution $g(x_i;\boldsymbol{\pi}^{(t)})_j:=\mathbb{P}(Y_t=j|X_t=x_i,\boldsymbol{\pi}^{(t)})$:

$$g(x_i;\boldsymbol{\pi}^{(t)})_j=\frac{\frac{\pi_j^{(t)}}{c_j}f(x_i)_j}{\sum_{l=1}^K\frac{\pi_l^{(t)}}{c_l}f(x_i)_l} \tag{41}$$

E-step Construct $Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})$ as:

$$\begin{aligned}Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})&=\mathop{\mathbb{E}}_{\mathbb{Y}|\mathbb{X},\boldsymbol{\pi}^{(t)}}[\log\mathbb{P}(\boldsymbol{\pi}|\mathbb{X},\mathbb{Y},\boldsymbol{\alpha})]\\&=\sum_{i=1}^N\sum_{j=1}^K g(x_i;\boldsymbol{\pi}^{(t)})_j\log\pi_j+\sum_{l=1}^K(\alpha_l-1)\log\pi_l+Const\end{aligned} \tag{42}$$

M-step Maximize $Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})$ w.r.t $\boldsymbol{\pi}\in\Delta^{K-1}$:

$$\pi_j^{(t+1)}=\left(\arg\max_{\boldsymbol{\pi}\in\Delta^{K-1}} Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)})\right)_j=\frac{\sum_{i=1}^N g(x_i;\boldsymbol{\pi}^{(t)})_j+\alpha_j-1}{N+\sum_{l=1}^K(\alpha_l-1)} \tag{43}$$

end for

Output: $\mathbb{P}(Y_t=\cdot)=\boldsymbol{\pi}^{(T+1)}$

C. Model Performance Summary

Label Shift Estimation Error: The experimental result demonstrates that, in terms of label shift estimation error, our MAPLS-APL model generally outperform previous models on large scale datasets. As shown in Tab. 7, in terms of label shift estimation error, for each train-test label shift setting, our model outperform other models at least half of the time.

Train Set \ Test Set	Ordered LT	Shuffled LT	Dirichlet
CIFAR100/CIFAR100-LT	55% (Tab. 15, 17)	50% (Tab. 19)	52% (Tab. 21)
ImageNet/ImageNet-LT	82% (Tab. 23, 25)	90% (Tab. 27)	75% (Tab. 29)
Places/Places-LT	68% (Tab. 31, 33)	90% (Tab. 35)	92% (Tab. 37)

Table 7. **Label Shift Estimation performance summary.** The percentage of settings that our MAPLS-APL model outperform other label shift estimation models in terms of $(w - \hat{w})^2/K$ and the reference to original results.

Our MAPLS-APL model tends to outperform other model when source and target domain have highly imbalance label distribution.

Top1 Accuracy: Label shift estimation models generally do not compare their performance in terms of Top1 Accuracy [1,2,10,22]. In this work, we provide this metric to analyze if a label shift estimation model can be used to generally improve the accuracy of a classifier under label shift.

Train Set \ Test Set	Ordered LT	Shuffled LT	Dirichlet
CIFAR100/CIFAR100-LT	56% (Tab. 16, 18)	58% (Tab. 20)	58% (Tab. 22)
ImageNet/ImageNet-LT	59% (Tab. 24, 26)	80% (Tab. 28)	58% (Tab. 30)
Places/Places-LT	59% (Tab. 32, 34)	80% (Tab. 36)	75% (Tab. 38)

Table 8. **Label Shift Estimation and Correction Top1 Acc summary.** The percentage of settings that our MAPLS-APL model outperform other label shift estimation models and the baseline in terms of Top1 Accuracy and the reference to the original results.

Training time: We also analyze the average training time of each label shift estimation models on different datasets. Results are provided in Tab. 9. Our MAPLS-APL model have comparable training time with SOTA models. Our PSLS-APL model utilize the entire posterior at the cost of longer training time.

Model	CIFAR100	ImageNet	Places
MLLS	< 1	~ 20	~ 10
BBSE	< 1	~ 10	~ 2
RLLS	< 1	~ 200	~ 20
MAPLS-APL (**ours**)	< 1	~ 25	~ 10
PSLS-APL (**ours**)	~ 100	~ 1000	~ 200

Table 9. **Average training time comparison (seconds).** Training time of SOTA models and our MAPLS-APL/PSLS-APL models on different dataset on a single NVIDIA RTX 2080Ti GPU. Our PSLS-APL model trade longer training time for entire posterior information.

Remark on PSLS-APL model Unlike SOTA models and our MAPLS-APL model that provide point estimate of $\boldsymbol{\pi}$, our posterior sampling method PSLS-APL obtain i.i.d samples from the posterior $\mathbb{P}(\boldsymbol{\pi}|\mathbb{X}, \boldsymbol{\alpha})$. Thus PSLS-APL is not feasible to compare Top1 Accuracy with other models in Tab. 6.

D. More on Adaptive Prior Learning Model

D.1. More about Model Misspecification and Sampling Error

The left figure in Fig. 2 demonstrates how model misspectifcation can influence the performance of label shift estimation models. The results are obtained on the CIFAR100-100-LT dataset, with different classifiers f (trained on different CIFAR100-LT datasets) and fixed source label distribution $\mathbb{P}(Y_s = \cdot)$ and target label distribution $\mathbb{P}(Y_t = \cdot)$.

The right figure in Fig. 2 demonstrates that the relative sampling error of a label shift estimation model can be amplified when there is a large label shift between source and target label distribution. The results are obtained on the ImageNet-LT dataset, with different target label distributions $\mathbb{P}(Y_t = \cdot)$ (different Ordered LT target domain shift) and fixed source label distribution $\mathbb{P}(Y_s = \cdot)$ and classifier f.

D.2. Empirical Justification of APL model

With empirical evidence, we show that the best choice of the parameter for the Dirichlet prior $\boldsymbol{\alpha}$ or parameter $\lambda = N/(N + K(\alpha_0 - 1))$ in our MAPLS model is different with different source and target label distribution settings. Further, our proposed Adaptive Prior Learning model can give a good choice of λ.

We compared the performance of the MAPLS-APL model with MAPLS models that have fixed λ. The experiment is carried out on the ImageNet-LT datset with a uniform test set. As shown in Eq. (12), our MAPLS model degenerates to MLLS in the $\lambda = 1$ setting. As can be seen from Tab. 10, the best choice of λ varies with different target label distribution. MLLS generally performs worse than every MAPLS model. Our heuristic can give the λ that is close to the best choice.

Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-soft	48.6621	42.5194	40.3712	58.1909	73.8265	83.7752	86.1056	161.6308	163.4361	252.6892	291.1914
MAPLS-soft ($\lambda = 0.9$)	**12.9288**	**12.3030**	13.2342	15.0942	14.6698	18.6681	20.2939	45.5180	52.1702	98.6632	133.5647
MAPLS-soft ($\lambda = 0.7$)	15.8842	14.6270	**12.7211**	**10.3413**	6.1992	4.6816	4.4100	12.4146	22.5194	**51.8191**	**82.3434**
MAPLS-soft ($\lambda = 0.5$)	20.8323	18.7681	15.2712	11.3622	**5.5232**	2.4841	1.7870	8.8930	**21.6946**	52.7513	86.8700
MAPLS-soft ($\lambda = 0.3$)	24.9417	22.2047	17.5530	12.6299	5.6351	1.9094	1.1774	**9.1997**	24.3382	59.6757	98.3755
MAPLS-soft ($\lambda = 0.1$)	28.1548	24.9087	19.4085	13.7481	5.9058	**1.7629**	**1.1065**	10.2160	27.2829	66.6140	109.3945
MAPLS-APL-soft (**ours**)	13.8236	13.3801	12.5594	10.3940	5.8088	3.5968	3.6781	13.4849	24.4035	57.7872	90.3704

Table 10. Performance of $(w - \hat{w})^2/K$ on ImageNet-LT dataset, with Ordered Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$ and forward and backward order. Best among fixed λ models are in bold face. Each reported values are averaging over 10 different shuffled and random sampled test set.

E. Experiment Details

E.1. Classifiers Details

We implement the Neural Netwrok classifier models using PyTorch [27]. We train a ResNet32 [18] classifier for CIFAR100 and every CIFAR100-LT dataset with weight decay $5e^{-4}$ for 200 epochs. The learning rate is initialized at 0.1 and drops by a factor of 10 at epochs 100 and 150. For ImageNet and Places, we use the pre-trained ResNet50 and ResNet152 respectively a classifier. For ImageNet-LT and Places-LT, we train a ResNet50 [14] classifier with weight decay $2e^{-4}$ for 100 epochs. The learning rate is initialized to 0.1 and drops by a factor of 10 at epochs 60 and 80.

Dataset	Model	Setup	lr	weight decay	epoch	scheduler	mixup α
CIFAR100/CIFAR100-LT	ResNet32	Train from Scratch	0.1	$5e^{-4}$	200	lr decay 0.1 at $[100, 150]$	0.2
ImageNet	ResNet50	Pre-Trained on ImageNet	-	-	-	-	-
ImageNet-LT	ResNet50	Train from Scratch	0.1	$2e^{-4}$	100	lr decay 0.1 at $[60, 80]$	0.2
Places	ResNet152	Pre-Trained on Places	-	-	-	-	-
Places-LT	ResNet152	Pre-Trained on ImageNet	0.001	$1e^{-4}$	100	lr decay 0.1 at $[60, 80]$	0.1

Table 11. Neural Network classifier setup used in our model.

For all the models training from scratch, we apply MixUp [41] with the parameter set to 0.2 during training. This is because MixUp is known to help increase Neural Network classifiers' calibration performance [35] and MLLS works better when classifier is calibrated on the source domain [1, 10].

E.2. Label Shift Estimation Models Details

We report the performance of previous methods based on the source code below. MLLS code is provided by Alexandari *et al*. [1] which have included the source code of RLLS [2] and BBSE [22] with their original github page provided in the Tab. 12. Only RLLS has hyperparameter in their model. We follow Alexandari *et al*. [1] and RLLS original implementation to set the hyperparameter to be $\alpha = 0.01$.

Model Name	Source Code	Date of Retrieval
MLLS [1, 32]	https://github.com/kundajelab/labelshiftexperiments https://github.com/kundajelab/abstention	Aug 2022 Aug 2022
BBSE [22]	https://github.com/flaviovdf/label-shift	Aug 2022
RLLS [2]	https://github.com/Angie-Liu/labelshift	Aug 2022

Table 12. Source Code details of reproduced existing label shift estimation models.

For MLLS and our proposed model, we follow MLLS [1] to initialize target label distribution the same as source domain label distribution $\boldsymbol{\pi}^{(0)} = \mathbf{c}$. Each EM have $T = 100$ iteration to get the final estimation. As shown in Fig. 6, MLLS and our MAPLS-APL algorithm have converged after 100 iteration.

Note that confusion matrix based method like BBSE requires a validation set from source domain P_s to construct $K \times K$ confusion matrix. However, this is not feasible for dataset with larget K. For ImageNet, this means estimating $1000 \times 1000 = 1e^6$ elements in confusion matrix with only $5e^4$ validation samples. Therefore in our experiments, we use train set data to estimating the $K \times K$ confusion matrix. In ImageNet case, the $1e^6$ element in confusion matrix is then estimated with $1.28e^6$ samples rather than just $5e^4$ samples. This approach may introduce extra error, but is more feasible in practice [22].

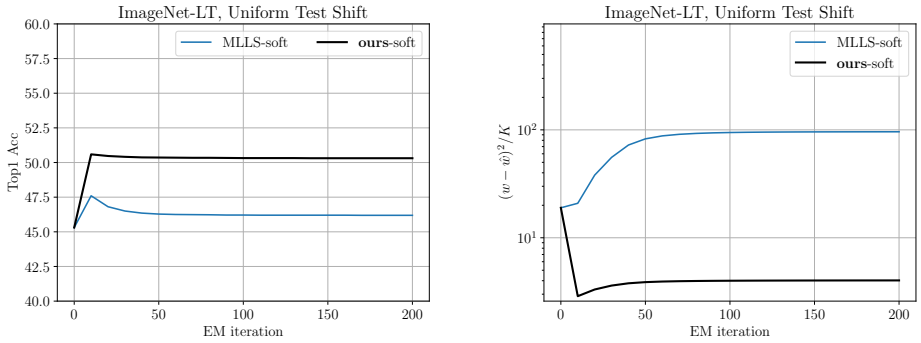

Figure 6. Performance of EM algorithm MLLS-soft and our MAPLS-APL-soft with different number of EM iterations, on ImageNet-LT dataset with uniform test set. Each algorithm have converged after 100 iteration.

E.3. Experiment Setup Details

Label Shift Estimation Error: For label shift estimation error $(w - \hat{w})^2/K$, BBSE and RLLS are able to directly output $\hat{w}$ as a prediction to ground truth $w = \mathbb{P}(Y_t = \cdot)/\mathbb{P}(Y_s = \cdot)$. Therefore their performance can be computed directly. MLLS and our MAPLS/MAPLS-APL model predicts ground truth $\boldsymbol{\pi}$ with $\hat{\boldsymbol{\pi}}$. Thus we follow MLLS to compute $\hat{w} = \hat{\boldsymbol{\pi}}/\mathbf{c}$, where $\mathbf{c}$ is the source label distribution estimated by MLE given source domain labelled data.

Top1 Accuracy: The Top1 Accuracy of each label shift estimation model is obtained by first estimating the label shift with corresponding model, then correct label shift on target domain for classifier f, with offline label shift correction method defined in Eq. (5). We also report Top1 Accuracy on baseline classifier without any label shift correction.

Train and Test Sets: We test our MAPLS/MAPLS-APL model with as many different label shift settings as we can. Our experiments includes all the train-test set combinations of a train set in Tab. 13 and a test set in Tab. 14.

Dataset	Setup	Imbalance Ratio	Data Size	# of Classes	Top class sample	Tail class sample
CIFAR100 [21]	Original	None	50k	100	500	500
	Long-tailed	2	36.0k	100	500	250
	Long-tailed	5	24.8k	100	500	100
	Long-tailed	10	19.5k	100	500	50
	Long-tailed	20	15.9k	100	500	25
	Long-tailed	50	12.6k	100	500	10
	Long-tailed	100	10.8k	100	500	5
	Long-tailed	200	9.5k	100	500	2
Places [44]	Original	None	1803.4k	365	5000	3068
	Long-tailed	996	62.5k	365	4980	5
ImageNet [31]	Original	None	1281.1k	1000	1300	732
	Long-tailed	256	115.8k	1000	1280	5

Table 13. Detailed information of train sets with different label shift in our paper.

Target Shift	Parameters
Original (Uniform)	None
Ordered Long-Tail [16]	$R = \{2, 5, 10, 50\}$, Order = "Forward", "Backward"
Shuffled Long-Tail	$R = \{2, 5, 10, 50\}$
Dirichlet [22]	$\alpha = 1.0, 10$

Table 14. Detailed information of test sets with different label shift in our paper.

F. CIFAR100 and CIFAR100-LT dataset results

F.1. Ordered Long-Tailed test set

CIFAR100											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	0.1112	0.1088	0.0785	0.0703	0.0534	0.0443	0.0433	0.0462	0.0483	0.0543	0.0577
MLLS-soft	**0.0519**	**0.0534**	**0.0368**	0.0325	0.0256	0.0199	0.0224	0.0303	**0.0317**	**0.0398**	**0.0436**
BBSE-hard	0.1871	0.1388	0.0776	0.0416	0.0154	0.0089	0.0184	0.0482	0.0855	0.1411	0.1892
BBSE-soft	0.0984	0.0766	0.0453	0.0285	**0.0142**	0.0102	**0.0153**	0.0325	0.0494	0.0771	0.1017
RLLS-hard	1.0940	0.7798	0.4255	0.2150	0.0411	0.0000	0.0411	0.2150	0.4255	0.7798	1.0940
RLLS-soft	1.0939	0.7798	0.4255	0.2150	0.0411	**0.0000**	0.0411	0.2150	0.4255	0.7798	1.0939
MAPLS-APL-hard (**Ours**)	0.1149	0.1003	0.0662	0.0520	0.0349	0.0286	0.0305	0.0403	0.0505	0.0679	0.0799
MAPLS-APL-soft (**Ours**)	0.0738	0.0638	0.0385	**0.0271**	0.0164	**0.0115**	0.0155	**0.0272**	0.0352	0.0517	0.0633
CIFAR100-2-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	0.0584	0.0519	0.0576	0.0605	0.0585	0.0579	0.0654	0.0778	0.0842	0.0927	0.1166
MLLS-soft	**0.0462**	**0.0398**	**0.0377**	0.0368	0.0323	0.0283	0.0343	0.0474	0.0599	0.0761	0.0881
BBSE-hard	0.1477	0.1074	0.0703	0.0456	0.0197	0.0128	0.0217	0.0625	0.1075	0.2009	0.2638
BBSE-soft	0.0727	0.0543	0.0379	**0.0297**	0.0187	0.0137	0.0179	0.0352	0.0521	0.0887	0.1107
RLLS-hard	0.6031	0.4099	0.1908	0.0680	0.0000	0.0458	0.1946	0.5633	0.9705	1.6457	2.2487
RLLS-soft	0.6031	0.4099	0.1908	0.0680	**0.0000**	0.0458	0.1946	0.5633	0.9705	1.6457	2.2487
MAPLS-APL-hard (**Ours**)	0.0737	0.0598	0.0508	0.0462	0.0322	0.0210	0.0356	0.0628	0.0818	0.1150	0.1567
MAPLS-APL-soft (**Ours**)	0.0822	0.0674	0.0529	0.0455	0.0289	**0.0086**	**0.0130**	**0.0274**	**0.0390**	**0.0627**	**0.0806**
CIFAR100-5-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	0.0816	0.0793	0.1073	0.1216	0.1561	0.1853	0.2457	0.3504	0.4258	0.5287	0.6359
MLLS-soft	**0.0775**	**0.0748**	0.0804	0.0995	0.0826	0.0942	0.1211	0.1862	0.1767	**0.3124**	0.3744
BBSE-hard	0.1881	0.1816	0.1164	0.0984	0.0718	0.0377	0.0671	0.1560	0.2621	0.6362	0.7353
BBSE-soft	0.0841	0.0713	**0.0570**	0.0537	**0.0356**	0.0408	0.0574	0.1261	0.1708	0.3669	0.4281
RLLS-hard	0.3220	0.1866	0.0447	**0.0001**	0.1076	0.3862	0.8978	1.9947	3.1599	5.0756	6.7923
RLLS-soft	0.3220	0.1866	0.0447	0.0001	0.1076	0.3862	0.8977	1.9947	3.1598	5.0756	6.7922
MAPLS-APL-hard (**Ours**)	0.1204	0.1025	0.0950	0.0807	0.0481	**0.0249**	0.0919	0.2702	0.4217	0.6506	0.8226
MAPLS-APL-soft (**Ours**)	0.2004	0.1974	0.1847	0.2024	0.1415	0.0600	**0.0351**	**0.0867**	**0.1390**	0.3252	**0.2978**
CIFAR100-10-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	**0.0643**	**0.0814**	0.1076	0.1646	0.2924	0.4731	0.7737	1.3339	1.8762	2.7663	3.4112
MLLS-soft	0.2085	0.2307	0.3143	0.3183	0.3521	0.4380	0.3576	1.2786	0.9244	1.2372	2.6438
BBSE-hard	0.3988	0.6186	0.6794	0.3136	1.0622	0.1699	0.2268	0.8856	0.8115	1.1396	1.7270
BBSE-soft	0.1796	0.1941	0.1825	0.1526	0.2604	0.2252	0.2897	0.8663	**0.7290**	**0.9373**	**1.5710**
RLLS-hard	0.2006	0.0829	0.0003	0.0693	0.5182	1.2920	2.5931	5.2848	8.1105	12.7453	16.9049
RLLS-soft	0.2006	0.0829	**0.0003**	0.0693	0.5182	1.2920	2.5931	5.2848	8.1105	12.7452	16.9047
MAPLS-APL-hard (**Ours**)	0.1688	0.1465	0.1021	**0.0666**	**0.0288**	**0.0478**	0.3065	1.2248	2.2082	3.7278	4.8302
MAPLS-APL-soft (**Ours**)	0.6830	0.7482	0.9220	0.9069	0.7311	0.3466	**0.1153**	**0.4548**	0.8022	1.2851	2.1746

Table 15. Performance of label shift estimation error $(w - \hat{w})^2/K$ on CIFAR100 and CIFAR100-LT dataset, with imbalance ratio $R = \{2, 5, 10\}$. Test set have ordered Long-Tailed shift. Each reported values are averaging over 20 different shuffled and random sampled test set.

CIFAR100-20-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	**0.1587**	0.2265	0.3015	0.5228	0.9652	1.5080	2.3236	3.9671	5.5135	8.0038	10.2356
MLLS-soft	0.6706	0.8996	1.2976	1.7197	2.0114	2.3082	3.8745	4.2966	3.3055	5.3265	9.4385
BBSE-hard	1.2515	2.0845	5.0574	3.4216	1.3311	26.4551	3.6669	10.0455	3.4961	9.4552	**5.4054**
BBSE-soft	0.9353	0.8459	1.2806	1.2914	1.0199	1.6793	1.9814	3.4160	4.5680	7.5762	6.9400
RLLS-hard	0.0997	0.0110	0.0807	0.4611	1.8463	3.9698	7.4201	14.4547	21.8131	33.8971	44.7735
RLLS-soft	0.0997	**0.0110**	**0.0807**	0.4611	1.8463	3.9702	7.4202	14.4550	21.8130	33.8967	44.7730
MAPLS-APL-hard (**Ours**)	0.4939	0.4087	0.3151	**0.2506**	**0.1197**	**0.1526**	**0.7543**	3.0308	5.6868	9.9517	13.1944
MAPLS-APL-soft (**Ours**)	2.1248	2.5210	3.1323	3.3225	2.7560	1.3423	0.8613	**1.0474**	**1.5530**	**3.2712**	6.3812

CIFAR100-50-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	0.4564	0.7598	1.5995	2.9962	6.1775	10.6642	17.3656	30.3777	44.2637	66.0673	85.3506
MLLS-soft	3.6399	4.6052	4.2592	5.2411	10.2603	10.5574	20.0799	38.2874	32.9386	75.1315	78.7784
BBSE-hard	91.1766	$5.7e^6$	$5.4e^5$	407.9134	$4.4e^6$	$2.0e^7$	$4.5e^8$	$1.0e^7$	$7.8e^6$	$6.7e^3$	$1.3e^7$
BBSE-soft	96.1977	38.2666	254.0946	5.8270	$1.6e^5$	306.8911	265.8220	880.7144	$8.5e^3$	$2.3e^4$	$1.5e^4$
RLLS-hard	**0.0054**	**0.0925**	0.9226	2.7620	8.4237	16.6034	29.9977	56.5451	84.8068	131.7854	173.7755
RLLS-soft	0.0054	0.0906	**0.9193**	2.7622	8.4295	16.6977	29.9299	56.8225	85.0470	131.6754	173.7434
MAPLS-APL-hard (**Ours**)	1.5923	1.4350	1.0658	**0.6335**	**0.3920**	**1.2321**	**4.5314**	15.2032	29.4610	54.8555	78.7368
MAPLS-APL-soft (**Ours**)	14.4861	15.3854	16.4496	13.9173	16.4473	9.6150	4.9813	**8.2935**	**10.1994**	**22.8203**	**47.0972**

CIFAR100-100-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	44.9205	38.0887	45.8544	42.3600	43.7414	47.8233	63.87	103.37	149.87	220.80	273.28
MLLS-soft	3.0637	6.0161	10.0665	11.1877	28.0022	41.4067	54.1261	101.44	161.84	227.24	317.32
BBSE-hard	$2.3e^4$	$1.5e^4$	$3.1e^3$	$4.9e^4$	$9.7e^7$	$1.2e^8$	$3.9e^5$	$9.8e^6$	$1.5e^5$	$2.2e^7$	$2.4e^6$
BBSE-soft	153.7269	$3.2e^5$	$8.5e^3$	493.4652	$7.2e^4$	$9.1e^4$	$1.2e^5$	$5.3e^3$	$4.3e^4$	$1.5e^5$	$2.4e^4$
RLLS-hard	**0.0873**	0.7020	**3.5932**	**9.3034**	26.2131	50.6648	89.58	169.332	253.10	391.91	518.30
RLLS-soft	0.0870	**0.7008**	3.5943	9.3038	26.2526	50.6641	89.83	169.23	253.26	391.96	518.00
MAPLS-APL-hard (**Ours**)	20.0839	18.3336	14.7602	9.9220	**2.8452**	**0.8477**	**6.75**	**31.63**	**65.97**	**130.14**	**191.80**
MAPLS-APL-soft (**Ours**)	18.2872	21.4946	18.0630	13.0629	4.3482	1.8832	7.50	35.66	68.15	138.09	218.09

CIFAR100-200-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	39.4734	54.3263	119.9399	157.9984	271.7572	419.75	630.63	1012.82	1346.69	1931.56	2567.15
MLLS-soft	10.7235	11.2820	47.2927	55.0693	131.0271	232.43	350.49	597.74	902.51	1386.22	1804.34
BBSE-hard	$3.2e^5$	$9.3e^6$	$6.5e^6$	$8.9e^8$	$6.8e^4$	$2.6e^9$	$6.5e^6$	$1.0e^4$	$2.6e^4$	$1.2e^5$	$1.9e^5$
BBSE-soft	$1.4e^5$	$2.3e^5$	$2.3e^4$	$7.6e^6$	772.48	$1.3e^5$	$1.1e^5$	8783.57	$1.4e^5$	$1.0e^4$	$1.2e^9$
RLLS-hard	**0.8851**	**3.5660**	**14.4068**	35.0877	97.2357	187.63	334.97	642.61	972.67	1532.69	2051.74
RLLS-soft	0.8875	3.5685	14.4299	35.0986	97.2723	187.64	335.00	642.44	972.67	1532.79	2051.69
MAPLS-APL-hard (**Ours**)	50.3299	48.6633	40.0183	**25.6926**	11.3733	21.34	72.55	233.90	440.66	839.90	1249.41
MAPLS-APL-soft (**Ours**)	70.7102	64.3605	56.5578	35.5457	**7.3678**	**8.33**	**49.87**	**198.37**	**376.73**	**742.34**	**1128.60**

Table 16. Performance of label shift estimation error $(w - \hat{w})^2/K$ on CIFAR100-LT dataset, with imbalance ratio $R = \{20, 50, 100, 200\}$. Test set have ordered Long-Tailed shift. Each reported values are averaging over 20 different shuffled and random sampled test set.

CIFAR100											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	69.05	69.01	69.81	69.88	70.28	70.53	70.75	71.00	70.95	70.96	71.28
MLLS-hard	72.05	70.58	70.36	69.58	69.56	69.92	70.40	71.60	72.44	73.60	75.10
MLLS-soft	72.59	70.91	70.79	69.93	69.75	70.14	70.67	71.79	72.70	73.95	75.21
BBSE-hard	72.19	70.88	70.68	**70.19**	**69.93**	70.25	70.68	71.68	72.44	73.39	74.34
BBSE-soft	72.55	**71.19**	70.84	70.12	69.90	70.24	70.75	**71.90**	72.66	73.78	74.82
RLLS-hard	69.05	69.01	69.81	69.88	70.28	70.53	70.75	71.00	70.95	70.96	71.28
RLLS-soft	69.05	69.01	69.81	69.88	70.28	**70.53**	70.75	71.00	70.95	70.96	71.28
MAPLS-APL-hard (**Ours**)	72.06	70.84	70.60	69.84	69.72	70.02	70.53	71.70	72.56	73.72	75.04
MAPLS-APL-soft (**Ours**)	**72.62**	71.17	**70.94**	70.07	69.89	70.28	70.75	71.89	**72.81**	**74.00**	**75.25**

CIFAR100-2-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	70.73	70.31	69.70	69.29	68.85	68.01	67.45	66.92	66.15	65.72	65.68
MLLS-hard	73.38	71.58	69.56	68.63	67.93	67.35	67.31	68.18	68.69	70.33	72.02
MLLS-soft	**73.81**	**72.03**	70.14	69.20	68.51	67.88	67.75	68.62	69.02	70.67	72.41
BBSE-hard	72.86	71.62	70.02	69.29	68.66	68.07	68.06	68.74	69.09	70.19	71.83
BBSE-soft	73.44	71.97	70.18	69.40	68.65	68.03	67.91	68.75	69.14	70.70	72.37
RLLS-hard	70.73	70.31	69.70	69.29	68.85	68.01	67.45	66.92	66.15	65.72	65.68
RLLS-soft	70.73	70.31	69.70	69.29	**68.85**	68.01	67.45	66.92	66.15	65.72	65.68
MAPLS-APL-hard (**Ours**)	73.37	71.71	69.87	68.96	68.31	67.88	67.65	68.54	68.87	70.62	72.21
MAPLS-APL-soft (**Ours**)	73.55	71.99	**70.22**	**69.41**	68.75	**68.19**	**68.13**	**68.88**	**69.26**	**71.01**	**72.61**

CIFAR100-5-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	70.64	69.75	68.03	66.62	64.37	62.59	60.80	58.25	56.74	55.04	53.10
MLLS-hard	71.44	69.59	66.99	65.41	63.49	62.35	61.79	61.38	62.27	63.42	64.31
MLLS-soft	71.75	69.85	67.43	66.01	64.21	63.42	62.81	62.63	63.58	**64.76**	**65.56**
BBSE-hard	71.13	69.56	67.37	66.08	64.24	63.55	62.97	62.47	63.29	63.88	64.45
BBSE-soft	**71.77**	**70.09**	67.68	66.27	64.37	63.46	62.86	62.55	63.39	64.57	64.84
RLLS-hard	70.64	69.75	68.03	66.62	64.37	62.59	60.80	58.25	56.74	55.04	53.10
RLLS-soft	70.64	69.75	**68.03**	**66.62**	64.37	62.59	60.80	58.25	56.74	55.04	53.10
MAPLS-APL-hard (**Ours**)	71.46	69.73	67.22	65.87	64.15	63.42	62.57	61.89	62.54	63.31	63.87
MAPLS-APL-soft (**Ours**)	71.32	69.58	67.34	66.00	**64.59**	**63.94**	**63.46**	**63.01**	**63.59**	64.59	65.23

CIFAR100-10-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	67.18	65.62	63.12	60.63	57.50	54.75	52.14	48.50	45.61	43.16	40.91
MLLS-hard	**67.51**	65.58	62.64	60.29	57.82	56.01	54.70	53.29	52.40	52.98	53.11
MLLS-soft	67.24	65.38	62.60	60.31	57.91	56.57	55.70	54.11	**53.88**	**55.30**	**55.38**
BBSE-hard	66.86	65.10	62.33	60.31	57.85	56.58	55.50	54.32	53.40	54.84	55.25
BBSE-soft	67.33	65.48	62.82	60.43	58.04	56.69	55.64	54.10	53.55	54.89	55.53
RLLS-hard	67.18	65.62	63.12	60.63	57.50	54.75	52.14	48.50	45.61	43.16	40.91
RLLS-soft	67.18	**65.62**	**63.12**	**60.63**	57.50	54.75	52.14	48.50	45.61	43.16	40.91
MAPLS-APL-hard (**Ours**)	67.33	65.46	62.75	60.52	58.29	56.83	55.49	53.71	52.48	52.22	52.00
MAPLS-APL-soft (**Ours**)	66.36	64.67	62.24	60.17	**58.19**	**56.98**	**55.78**	**54.44**	53.37	54.24	54.30

Table 17. Top1 Accuracy Performance on CIFAR100 and CIFAR100-LT dataset, with imbalance ratio $R = \{2, 5, 10\}$. Each reported values are averaging over 20 different shuffled and random sampled test set.

CIFAR100-20-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	65.65	63.02	59.88	57.05	52.71	49.11	45.58	41.27	38.26	34.73	32.47
MLLS-hard	64.98	62.15	59.14	56.45	52.97	50.49	48.58	47.05	46.33	45.67	45.36
MLLS-soft	64.42	61.45	58.59	56.52	53.45	51.63	49.97	48.72	**48.48**	**48.68**	**48.45**
BBSE-hard	63.65	61.23	58.16	55.75	53.67	49.06	49.71	48.25	47.79	48.43	48.46
BBSE-soft	64.54	61.72	58.76	56.61	53.70	51.40	49.86	48.56	47.79	48.07	48.35
RLLS-hard	65.65	63.02	59.88	57.05	52.71	49.11	45.58	41.27	38.26	34.73	32.47
RLLS-soft	**65.65**	**63.02**	**59.88**	57.05	52.71	49.11	45.58	41.27	38.26	34.73	32.47
MAPLS-APL-hard (**Ours**)	64.72	62.01	59.29	**57.14**	54.07	51.97	50.26	48.42	46.97	45.92	45.35
MAPLS-APL-soft (**Ours**)	63.65	60.98	58.43	56.44	**54.03**	**52.48**	**50.86**	**49.20**	48.35	48.04	47.51

CIFAR100-50-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	63.60	60.91	56.57	52.85	47.21	42.65	38.14	32.46	28.60	23.82	20.98
MLLS-hard	62.74	59.90	55.79	52.59	47.90	44.56	41.57	38.20	36.20	33.73	32.52
MLLS-soft	61.87	58.90	55.68	53.05	48.58	46.03	43.92	40.49	40.11	37.60	**37.53**
BBSE-hard	54.09	37.72	27.03	45.20	25.17	25.23	20.70	21.41	17.59	15.96	20.33
BBSE-soft	60.58	58.57	52.37	52.78	30.25	40.54	40.12	34.56	29.89	27.73	32.36
RLLS-hard	**63.60**	**60.90**	**56.57**	52.85	47.20	42.59	38.13	32.54	28.56	23.67	20.98
RLLS-soft	63.60	60.90	56.57	52.85	47.21	42.65	38.14	32.46	28.62	23.80	20.98
MAPLS-APL-hard (**Ours**)	62.31	59.67	56.28	**53.45**	**49.44**	46.52	43.94	40.60	38.54	35.56	34.17
MAPLS-APL-soft (**Ours**)	60.60	57.77	55.04	52.72	49.11	**46.98**	**44.91**	**41.84**	**40.99**	**38.12**	37.45

CIFAR100-100-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	67.28	64.02	58.48	53.87	47.18	41.76	36.24	29.35	25.00	19.91	16.40
MLLS-hard	64.52	61.48	56.67	53.02	48.53	45.09	41.77	37.98	36.03	34.63	33.94
MLLS-soft	65.64	62.55	58.22	54.43	49.62	46.07	42.47	38.80	36.36	35.27	33.93
BBSE-hard	41.42	34.08	42.00	33.25	16.48	23.44	21.76	18.68	23.48	14.65	17.58
BBSE-soft	63.97	28.54	40.39	48.33	25.90	20.57	24.02	30.18	24.94	26.02	24.12
RLLS-hard	**67.29**	**64.02**	**58.48**	53.87	47.18	41.76	36.22	29.37	25.00	19.91	16.40
RLLS-soft	67.28	64.02	58.48	53.87	47.17	41.76	36.24	29.36	25.02	19.91	16.43
MAPLS-APL-hard (**Ours**)	64.98	62.04	58.17	55.13	51.26	48.10	45.03	41.38	39.35	37.20	35.69
MAPLS-APL-soft (**Ours**)	64.88	61.57	58.03	**55.21**	**51.62**	**48.57**	**45.58**	**42.12**	**39.83**	**37.69**	**35.96**

CIFAR100-200-LT											
Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	63.82	60.19	54.47	49.59	42.21	36.59	31.01	24.13	19.33	14.48	11.38
MLLS-hard	60.72	57.24	52.34	48.32	41.69	37.22	32.97	27.94	25.63	23.56	22.05
MLLS-soft	61.68	58.94	53.32	49.48	43.39	39.37	35.05	30.31	27.05	25.00	22.76
BBSE-hard	35.57	26.79	26.55	23.58	28.65	18.59	18.21	22.53	21.31	18.24	16.84
BBSE-soft	50.93	30.32	46.57	19.40	38.04	22.80	22.96	23.37	18.04	21.16	13.43
RLLS-hard	63.82	60.19	54.47	49.59	42.21	36.59	31.00	24.13	19.33	14.48	11.36
RLLS-soft	**63.82**	**60.19**	54.47	49.59	42.21	36.59	31.01	24.13	19.33	14.48	11.38
MAPLS-APL-hard (**Ours**)	61.56	58.72	**54.69**	51.10	45.64	41.75	37.67	32.54	29.12	25.30	23.28
MAPLS-APL-soft (**Ours**)	61.71	59.04	54.60	**51.11**	**46.37**	**42.81**	**38.70**	**33.94**	**30.72**	**27.16**	**23.90**

Table 18. Top1 Accuracy Performance on CIFAR100-LT dataset, with imbalance ratio $R = \{20, 50, 100, 200\}$. Test set have ordered Long-Tailed shift.. Each reported values are averaging over 20 different shuffled and random sampled test set.

F.2. Shuffled Long-Tailed test set

Dataset	CIFAR100					CIFAR100-2-LT					CIFAR100-5-LT				
Test imbalance ratio	50	25	10	5	2	50	25	10	5	2	50	25	10	5	2
MLLS-hard	0.0689	0.0612	0.0572	0.0525	0.0490	0.0943	0.0802	0.0837	0.0699	0.0624	0.2704	0.2438	0.2682	0.2096	0.1975
MLLS-soft	**0.0445**	**0.0376**	**0.0339**	0.0293	0.0256	**0.0752**	**0.0578**	0.0503	0.0484	0.0349	0.2217	0.2229	0.1566	0.1847	0.1089
BBSE-hard	0.2046	0.1257	0.0846	0.0483	0.0179	0.1976	0.1277	0.0852	0.0484	0.0227	0.3985	0.2501	0.1877	0.1525	0.0623
BBSE-soft	0.1068	0.0677	0.0495	0.0306	**0.0157**	0.0983	0.0647	**0.0481**	**0.0329**	**0.0195**	**0.1958**	**0.1308**	**0.1130**	**0.0865**	**0.0519**
RLLS-hard	1.0940	0.6902	0.4255	0.2150	0.0411	1.2680	0.8147	0.5304	0.2889	0.0928	2.4890	1.7722	1.2103	0.7971	0.4583
RLLS-soft	1.0939	0.6902	0.4255	0.2150	0.0411	1.2680	0.8146	0.5304	0.2889	0.0928	2.4890	1.7722	1.2103	0.7971	0.4583
MAPLS-APL-hard (**ours**)	0.0851	0.0623	0.0527	0.0414	0.0331	0.1132	0.0864	0.0744	0.0532	0.0333	0.3376	0.2693	0.2320	0.1378	0.0641
MAPLS-APL-soft (**ours**)	0.0675	0.0462	0.0366	**0.0258**	0.0168	0.0945	0.0680	0.0529	0.0410	0.0208	0.2527	0.2091	0.1707	0.1490	0.0853

Dataset	CIFAR100-10-LT					CIFAR100-20-LT					CIFAR100-50-LT				
Test imbalance ratio	50	25	10	5	2	50	25	10	5	2	50	25	10	5	2
MLLS-hard	1.0163	0.9764	0.7425	0.6067	0.5128	**2.9892**	2.8297	2.2574	1.7950	1.6122	23.03	15.79	14.31	13.10	10.78
MLLS-soft	0.7521	0.7404	0.6555	0.4225	0.3646	3.2989	2.1103	2.1952	1.6323	2.1817	29.63	20.58	11.85	13.12	15.48
BBSE-hard	0.7289	0.6706	0.4809	0.2653	0.2378	11.0902	4.4542	7.7656	2.2442	$4.1e^3$	$1.9e^3$	$3.7e^6$	$3.4e^3$	$3.7e^6$	$5.9e^4$
BBSE-soft	**0.5575**	**0.5481**	**0.4207**	**0.2640**	0.1517	6.5697	**1.5792**	**1.2312**	1.3583	1.6058	40.22	42.34	170.93	49.30	41.54
RLLS-hard	4.8606	4.1294	2.8207	2.0885	1.4624	11.8313	9.2090	7.2257	5.4889	4.2530	45.02	32.64	25.97	22.22	17.59
RLLS-soft	4.8606	4.1294	2.8207	2.0885	1.4624	11.8314	9.2091	7.2259	5.4889	4.2462	44.91	32.62	25.95	22.33	17.55
MAPLS-APL-hard (**ours**)	1.2782	1.1249	0.7031	0.4276	**0.1419**	3.4461	2.8919	1.8438	**1.1389**	**0.3678**	**19.98**	**11.74**	**8.26**	**5.11**	**1.93**
MAPLS-APL-soft (**ours**)	0.9630	0.8499	0.7405	0.5407	0.4442	3.2916	2.6013	2.2956	2.0995	1.5115	20.40	14.44	11.36	7.93	8.66

Dataset	CIFAR100-100-LT					CIFAR100-200-LT					-
Test imbalance ratio	50	25	10	5	2	50	25	10	5	2	
MLLS-hard	72.08	65.37	61.28	49.67	49.53	501.61	511.87	513.67	460.38	407.01	
MLLS-soft	60.36	71.17	51.32	37.32	57.24	355.83	309.85	326.21	254.41	246.17	
BBSE-hard	$1.5e^5$	$8.9e^5$	$8.8e^4$	$1.8e^5$	$4.6e^7$	$2.3e^6$	$1.0e^6$	$2.7e^6$	$2.5e^5$	$2.5e^4$	
BBSE-soft	477.58	$7.6e4$	$3.5e4$	$4.3e4$	487.35	$1.1e^6$	$8.2e^5$	$3.2e^6$	$6.4e^5$	$5.6e^5$	
RLLS-hard	102.20	99.08	81.88	62.23	53.06	397.01	330.37	309.49	243.86	192.48	
RLLS-soft	102.21	99.00	81.85	62.11	53.10	397.03	330.40	309.54	243.90	192.52	
MAPLS-APL-hard (**ours**)	**38.81**	**30.56**	**19.83**	10.94	**2.97**	210.92	160.89	124.92	72.82	28.54	
MAPLS-APL-soft (**ours**)	39.97	34.67	21.30	**10.87**	4.88	**192.82**	**140.31**	**107.24**	**49.81**	**14.66**	

Table 19. Performance of $(w - \hat{w})^2/K$ on CIFAR100 dataset and Long-Tailed CIFAR100 dataset with imbalance ratio $R = \{2, 5, 10, 20, 50, 100, 200\}$, with shuffled Long-Tailed test set that have imbalance ratio $R = \{50, 25, 10, 5, 2\}$. Each reported values are averaging over 50 different shuffled and random sampled test set.

Dataset	CIFAR100					CIFAR100-2-LT					CIFAR100-5-LT					CIFAR100-10-LT				
Test imbalance ratio	50	25	10	5	2	50	25	10	5	2	50	25	10	5	2	50	25	10	5	2
Baseline	70.13	70.97	70.33	70.67	**70.38**	68.36	68.03	67.93	68.07	68.08	62.39	61.92	62.38	62.54	62.50	55.02	53.68	54.39	55.01	54.66
MLLS-hard	74.12	73.22	71.58	71.01	69.76	72.17	70.27	68.96	68.27	67.54	68.01	65.61	64.44	63.58	62.43	61.05	58.32	57.67	57.30	56.11
MLLS-soft	**74.45**	73.46	71.90	71.22	70.10	**72.55**	70.82	69.49	68.79	68.13	**68.71**	66.40	65.34	64.34	63.35	50.66	49.36	47.83	46.69	45.80
BBSE-hard	73.54	73.16	71.68	71.17	70.15	72.02	70.47	69.41	68.87	68.25	67.76	66.14	65.26	64.27	63.47	60.98	58.55	58.12	57.75	56.68
BBSE-soft	74.11	73.42	71.85	71.34	70.17	72.48	70.82	69.58	68.92	68.27	68.50	**66.42**	**65.46**	64.45	63.48	**61.37**	**58.79**	**58.24**	**57.77**	56.66
RLLS-hard	70.13	70.97	70.33	70.67	70.38	68.36	68.03	67.93	68.07	68.08	62.39	61.92	62.38	62.54	62.50	55.02	53.68	54.39	55.01	54.66
RLLS-soft	70.13	70.97	70.33	70.67	70.38	68.36	68.03	67.93	68.07	68.08	62.39	61.92	62.38	62.54	62.50	55.02	53.68	54.39	55.01	54.66
MAPLS-APL-hard (**ours**)	74.22	73.36	71.76	71.14	69.94	72.28	70.51	69.24	68.61	67.98	67.80	65.80	64.83	64.03	63.31	60.59	58.08	57.78	57.60	56.70
MAPLS-APL-soft (**ours**)	74.41	**73.54**	**72.03**	**71.37**	70.20	72.53	**70.88**	**69.59**	**68.94**	**68.33**	68.37	66.41	65.36	**64.50**	**63.87**	60.74	58.48	57.89	57.73	**56.89**

Dataset	CIFAR100-20-LT					CIFAR100-50-LT					CIFAR100-100-LT					CIFAR100-200-LT				
Test imbalance ratio	50	25	10	5	2	50	25	10	5	2	50	25	10	5	2	50	25	10	5	2
Baseline	49.77	48.83	48.39	48.89	49.11	41.97	42.58	43.10	42.54	42.48	40.99	41.26	41.19	42.14	41.96	35.83	35.69	36.24	36.08	36.56
MLLS-hard	56.14	53.46	52.32	51.46	50.72	48.47	47.77	47.01	45.55	44.61	49.51	48.28	46.67	46.06	45.40	42.77	40.31	39.02	37.50	37.32
MLLS-soft	**57.06**	**54.39**	**53.22**	52.50	51.75	**49.81**	**49.49**	**48.91**	47.18	46.24	50.66	49.36	47.83	46.69	45.80	44.63	42.40	40.87	39.43	38.62
BBSE-hard	55.79	54.02	52.36	52.53	29.93	28.02	25.24	36.34	24.17	25.46	21.64	24.52	23.89	27.61	17.06	22.04	19.28	22.45	22.75	24.31
BBSE-soft	56.59	**54.39**	**53.22**	**52.60**	51.51	48.86	48.43	44.26	45.24	43.69	43.90	31.68	38.00	30.53	40.39	26.31	21.61	16.87	23.00	24.19
RLLS-hard	49.77	48.83	48.39	48.89	49.12	41.89	42.57	43.11	42.57	42.38	40.99	41.28	41.20	42.14	41.96	35.83	35.69	36.24	36.08	36.56
RLLS-soft	49.77	48.83	48.39	48.89	49.11	41.98	42.59	43.11	42.54	42.48	40.99	41.27	41.19	42.14	41.96	35.83	35.69	36.24	36.08	36.56
MAPLS-APL-hard (**ours**)	55.73	53.54	52.71	52.30	52.10	48.61	48.43	48.09	47.07	46.51	50.51	49.94	48.85	48.49	48.31	44.70	42.89	42.42	41.68	41.77
MAPLS-APL-soft (**ours**)	56.22	54.06	53.02	52.56	**52.43**	49.04	48.88	48.60	**47.45**	**47.08**	**51.02**	**50.55**	**49.12**	**48.97**	**48.66**	**45.83**	**43.95**	**43.41**	**42.68**	**42.44**

Table 20. Performance of Top1 Accuracy on CIFAR100 dataset and Long-Tailed CIFAR100 dataset with imbalance ratio $R = \{2, 5, 10, 20, 50, 100, 200\}$, with shuffled Long-Tailed test set that have imbalance ratio $R = \{50, 25, 10, 5, 2\}$. Each reported values are averaging over 50 different shuffled and random sampled test set.

F.3. Dirichlet Shifted test set

Dataset	CIFAR100						CIFAR100-2-LT						CIFAR100-5-LT					
Test Dirichlet α	$\alpha = 10$			$\alpha = 1$			$\alpha = 10$			$\alpha = 1$			$\alpha = 10$			$\alpha = 1$		
Test Sample No.	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500
MLLS-hard	0.0690	0.0532	0.0482	0.0702	0.0575	0.0576	0.0932	0.0689	0.0622	0.0946	0.0743	0.0726	0.2600	0.2111	0.2021	0.3109	0.2573	0.2506
MLLS-soft	0.0448	0.0290	0.0244	**0.0436**	**0.0307**	**0.0330**	0.0634	0.0468	0.0365	**0.0670**	**0.0542**	**0.0536**	0.2060	0.1478	0.1511	0.2337	0.2013	0.1640
BBSE-hard	0.0330	0.0281	0.0281	0.1792	0.1798	0.1744	0.0396	0.0332	0.0294	0.1813	0.1787	0.1754	0.0986	0.1011	0.0919	0.3597	0.3384	0.3033
BBSE-soft	**0.0290**	0.0212	0.0199	0.0950	0.0910	0.0890	**0.0383**	**0.0285**	**0.0223**	0.0894	0.0795	0.0786	**0.1034**	**0.0694**	**0.0717**	**0.1923**	**0.1658**	**0.1447**
RLLS-hard	0.1037	0.1024	0.1009	0.9902	1.0078	0.9966	0.1591	0.1627	0.1573	1.1288	1.1823	1.1718	0.6087	0.5633	0.5805	2.2550	2.1548	2.2429
RLLS-soft	0.1037	0.1024	0.1009	0.9902	1.0078	0.9966	0.1591	0.1627	0.1573	1.1288	1.1823	1.1718	0.6087	0.5632	0.5805	2.2550	2.1547	2.2429
MAPLS-APL-hard (**ours**)	0.0471	0.0372	0.0349	0.0808	0.0712	0.0713	0.0571	0.0447	0.0411	0.1091	0.0982	0.0970	0.1184	0.0981	0.1011	0.3486	0.3275	0.3000
MAPLS-APL-soft (**ours**)	0.0308	**0.0210**	**0.0188**	0.0612	0.0538	0.0542	0.0398	0.0320	0.0282	0.0859	0.0747	0.0757	0.1192	0.1193	0.1164	0.2405	0.2369	0.2088

Dataset	CIFAR100-10-LT						CIFAR100-20-LT						CIFAR100-50-LT					
Test Dirichlet α	$\alpha = 10$			$\alpha = 1$			$\alpha = 10$			$\alpha = 1$			$\alpha = 10$			$\alpha = 1$		
Test Sample No.	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500
MLLS-hard	0.6699	0.5618	0.5169	0.9042	0.9943	1.0304	2.0207	1.7062	1.6656	2.9207	3.0951	2.5477	11.93	11.36	11.70	22.20	19.33	17.51
MLLS-soft	0.6194	0.5405	0.3526	**0.7268**	0.6015	0.7655	2.8273	2.2490	2.8149	1.8841	**2.7385**	1.9305	24.31	14.75	15.56	23.05	20.14	21.29
BBSE-hard	0.3034	0.2696	0.7249	0.7961	0.7386	0.7871	13.4587	2.6115	3.8913	2.6597	4.0435	4.5142	$1.4e^7$	$5.5e^4$	$1.5e^5$	$1.7e^5$	$9.9e^4$	$2.2e^3$
BBSE-soft	0.4020	**0.2089**	0.2595	0.5639	**0.4615**	**0.6340**	5.6748	1.3563	2.4176	**1.5559**	2.8628	**1.5493**	$1.3e^3$	97.03	$4.4e^3$	$6.2e^2$	75.57	$2.4e^3$
RLLS-hard	1.7126	1.6543	1.6184	4.6657	4.6950	5.0918	4.8416	4.7709	4.6082	11.3887	11.7636	10.7253	18.70	18.59	19.25	41.60	36.61	36.57
RLLS-soft	1.7126	1.6543	1.6184	4.6657	4.6950	5.0918	4.8418	4.7710	4.6083	11.3887	11.7636	10.7253	18.72	18.65	19.14	41.59	36.61	36.56
MAPLS-APL-hard (**ours**)	**0.2939**	0.2600	**0.2455**	1.1216	1.2356	1.3374	**0.7200**	**0.6593**	**0.6125**	3.2722	3.6561	3.0976	**3.24**	**2.88**	**3.13**	18.09	**15.49**	**14.54**
MAPLS-APL-soft (**ours**)	0.4253	0.5086	0.4863	0.8567	0.8497	1.0113	1.6224	1.5438	1.8489	2.7654	3.1580	2.7714	10.04	9.30	8.34	**17.89**	17.03	19.55

Dataset	CIFAR100-100-LT						CIFAR100-200-LT						-					
Test Dirichlet α	$\alpha = 10$			$\alpha = 1$			$\alpha = 10$			$\alpha = 1$			-			-		
Test Sample No.	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500						
MLLS-hard	55.20	52.00	48.73	71.50	62.18	63.34	436.79	428.50	429.92	536.44	571.00	508.16						
MLLS-soft	51.97	45.55	42.43	68.44	61.42	62.14	254.46	255.35	232.31	351.04	370.53	349.11						
BBSE-hard	$3.2e^8$	$7.9e^6$	$1.6e^4$	$5.8e^6$	$4.8e^5$	$2.3e^4$	$2.3e^6$	$2.1e^6$	$2.3e^5$	$1.2e^5$	$2.7e^6$	$1.6e^6$						
BBSE-soft	$1.2e^4$	$2.3e^4$	$2.7e^3$	$2.5e^3$	$2.5e^3$	$3.4e^4$	$2.5e^4$	$1.3e^3$	$1.1e^5$	$3.1e^5$	$1.1e^4$	$5.7e^6$						
RLLS-hard	56.02	57.64	56.13	120.79	108.03	112.63	206.15	207.55	201.14	382.66	435.33	356.78						
RLLS-soft	55.96	57.71	56.13	120.71	108.00	112.60	206.10	207.57	201.18	382.68	435.29	356.77						
MAPLS-APL-hard (**ours**)	**5.47**	**5.53**	**5.48**	**41.47**	**38.81**	**36.87**	43.01	42.58	37.98	203.42	245.97	184.70						
MAPLS-APL-soft (**ours**)	6.93	7.46	7.09	45.26	39.51	41.07	**27.79**	**25.59**	**23.06**	**186.10**	**221.36**	**170.88**						

Table 21. Performance of label shift estimation error $(w - \hat{w})^2/K$ on CIFAR100 dataset and Long-Tailed CIFAR100 dataset with imbalance ratio $R = \{2, 5, 10, 20, 50, 100, 200\}$, with Dirichlet test set shift that have $\alpha = \{1, 10\}$ and total test sample number $\{2500, 5000, 7500\}$ in each setting. Each reported values are averaging over 50 different shuffle and random sampled test set.

Dataset	CIFAR100						CIFAR100-2-LT						CIFAR100-5-LT					
Test Dirichlet α	$\alpha = 10$			$\alpha = 1$			$\alpha = 10$			$\alpha = 1$			$\alpha = 10$			$\alpha = 1$		
Test Sample No.	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500
Baseline	**70.59**	70.41	70.44	70.73	70.80	70.20	67.86	67.96	68.01	67.96	68.00	68.19	62.67	62.55	62.55	62.16	62.42	62.71
MLLS-hard	69.98	70.06	70.21	73.97	74.34	73.78	67.24	67.68	67.85	71.20	71.53	71.75	62.68	62.72	62.94	66.52	67.09	67.19
MLLS-soft	70.33	70.44	70.52	**74.22**	**74.65**	**74.05**	67.76	68.20	68.33	**71.65**	**72.01**	**72.15**	63.51	63.50	63.79	**67.43**	**68.02**	**68.10**
BBSE-hard	70.53	70.53	70.53	73.59	73.80	73.17	68.08	68.37	68.39	71.15	71.34	71.45	63.88	63.70	63.88	66.76	67.25	67.24
BBSE-soft	70.56	**70.57**	70.61	74.01	74.27	73.64	68.05	68.36	68.45	71.51	71.78	71.96	63.81	63.72	63.91	67.22	67.74	67.82
RLLS-hard	70.59	70.41	70.44	70.73	70.80	70.20	67.86	67.96	68.01	67.96	68.00	68.19	62.67	62.55	62.55	62.16	62.42	62.71
RLLS-soft	70.59	70.41	70.44	70.73	70.80	70.20	67.86	67.96	68.01	67.96	68.00	68.19	62.67	62.55	62.55	62.16	62.42	62.71
MAPLS-APL-hard (**ours**)	70.28	70.28	70.39	74.01	74.24	73.69	67.74	68.03	68.14	71.21	71.49	71.71	63.57	63.47	63.59	66.45	66.93	67.04
MAPLS-APL-soft (**ours**)	70.54	70.56	**70.61**	74.20	74.51	73.88	**68.15**	**68.41**	**68.47**	71.62	71.87	72.05	**64.19**	**63.98**	**64.14**	67.11	67.62	67.67

Dataset	CIFAR100-10-LT						CIFAR100-20-LT						CIFAR100-50-LT					
Test Dirichlet α	$\alpha = 10$			$\alpha = 1$			$\alpha = 10$			$\alpha = 1$			$\alpha = 10$			$\alpha = 1$		
Test Sample No.	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500
Baseline	54.60	54.73	54.84	54.57	54.85	54.33	49.09	49.16	49.25	49.85	48.97	49.61	42.72	42.86	42.65	41.76	42.94	42.28
MLLS-hard	56.02	56.51	56.64	59.81	60.16	59.81	50.89	51.10	51.18	55.41	55.03	55.43	44.81	45.11	45.02	47.81	48.69	48.30
MLLS-soft	56.50	56.91	56.95	**60.30**	**60.72**	**60.35**	51.81	52.01	52.20	**56.43**	**56.04**	**56.42**	46.32	46.51	46.65	**49.61**	**50.58**	**50.17**
BBSE-hard	56.63	57.01	56.81	59.89	60.39	59.73	50.11	52.07	51.96	55.82	55.15	55.54	24.82	26.26	25.05	23.63	32.35	36.75
BBSE-soft	56.66	57.04	57.00	60.18	60.62	60.09	51.58	52.03	52.08	56.47	55.67	56.19	40.16	44.02	38.49	45.95	48.19	43.26
RLLS-hard	54.60	54.73	54.84	54.57	54.85	54.33	49.09	49.16	49.25	49.85	48.97	49.61	42.68	42.82	42.59	41.70	42.93	42.27
RLLS-soft	54.60	54.73	54.84	54.57	54.85	54.33	49.09	49.16	49.25	49.85	48.97	49.61	42.73	42.87	42.65	41.75	42.94	42.29
MAPLS-APL-hard (**ours**)	56.68	56.93	56.98	59.54	59.67	59.30	52.15	52.31	52.36	55.25	54.65	55.15	46.81	46.87	46.73	48.01	48.92	48.34
MAPLS-APL-soft (**ours**)	**56.92**	**57.11**	**57.02**	59.71	59.93	59.54	**52.63**	**52.64**	**52.64**	55.62	55.11	55.43	**47.35**	**47.29**	**47.15**	48.94	49.63	49.05

Dataset	CIFAR100-100-LT						CIFAR100-200-LT						-					
Test Dirichlet α	$\alpha = 10$			$\alpha = 1$			$\alpha = 10$			$\alpha = 1$			-			-		
Test Sample No.	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500						
Baseline	41.76	41.58	41.86	41.72	41.83	42.26	36.54	36.84	36.72	36.49	36.51	36.63						
MLLS-hard	44.69	45.12	45.46	49.46	49.98	50.10	37.17	37.80	37.85	41.98	41.88	42.25						
MLLS-soft	45.47	45.96	46.62	50.53	50.62	50.97	38.56	39.81	40.18	43.55	44.36	44.09						
BBSE-hard	23.57	29.21	26.14	18.09	25.43	25.86	20.50	21.44	26.67	24.96	21.26	19.18						
BBSE-soft	36.88	42.21	37.61	32.59	37.69	33.57	29.51	32.13	24.92	22.03	30.07	23.47						
RLLS-hard	41.71	41.58	41.86	41.76	41.83	42.25	36.53	36.81	36.72	36.49	36.49	36.63						
RLLS-soft	41.76	41.58	41.86	41.72	41.83	42.27	36.54	36.84	36.72	36.49	36.51	36.63						
MAPLS-APL-hard (**ours**)	47.91	48.04	48.30	50.41	50.89	51.13	41.67	42.13	42.05	44.29	43.91	44.45						
MAPLS-APL-soft (**ours**)	**48.27**	**48.44**	**48.85**	**50.89**	**51.18**	**51.52**	**42.54**	**43.23**	**43.18**	**45.20**	**45.30**	**45.52**						

Table 22. Performance of Top1 Accuracy on CIFAR100 dataset and Long-Tailed CIFAR100 dataset with imbalance ratio $R = \{2, 5, 10, 20, 50, 100, 200\}$, with Dirichlet test set shift that have $\alpha = \{1, 10\}$ and total test sample number $\{2500, 5000, 7500\}$ in each setting. Each reported values are averaging over 50 different shuffled and random sampled test set.

G. ImageNet and ImageNet-LT dataset results

G.1. Ordered Long-Tailed test set

Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	0.1166	0.1052	0.0880	0.0799	0.0708	0.0705	0.0862	0.1169	0.1483	0.1962	0.2463
MLLS-soft	0.1027	0.0927	0.0781	0.0690	0.0581	0.0575	0.0697	0.0962	0.1224	0.1744	0.2195
BBSE-hard	0.1171	0.0972	0.0783	0.0669	0.0556	0.0529	0.0638	0.0903	0.1194	0.1725	0.2191
BBSE-soft	0.1182	0.1010	0.0831	0.0727	0.0607	0.0600	0.0734	0.1037	0.1304	0.1850	0.2352
RLLS-hard	1.0846	0.7676	0.4097	0.2018	0.0350	0.0059	0.0692	0.3012	0.5848	1.0887	1.5532
RLLS-soft	1.0846	0.7676	0.4097	0.2018	0.0350	**0.0059**	0.0692	0.3012	0.5848	1.0887	1.5532
MAPLS-APL-hard (**ours**)	0.1068	0.0891	0.0664	0.0539	0.0422	0.0419	0.0572	0.0923	0.1298	0.1906	0.2508
MAPLS-APL-soft (**ours**)	**0.0942**	**0.0777**	**0.0579**	**0.0453**	**0.0319**	0.0296	**0.0430**	**0.0743**	**0.1064**	**0.1629**	**0.2149**

Table 23. Performance of $(w - \hat{w})^2/K$ on ImageNet dataset, with Ordered Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 20 different shuffled and random sampled test set.

Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	78.18	78.16	77.66	77.32	76.70	76.13	75.47	74.63	73.96	72.96	72.29
MLLS-hard	78.61	78.03	76.97	76.34	75.50	75.03	74.39	74.01	73.90	73.85	73.89
MLLS-soft	78.82	78.25	77.16	76.55	75.75	75.27	74.70	74.33	74.28	74.18	74.35
BBSE-hard	78.59	78.08	77.17	76.56	75.85	75.33	74.74	74.27	74.16	73.93	73.99
BBSE-soft	78.53	77.94	77.01	76.51	75.70	75.22	74.64	74.27	74.15	73.90	73.93
RLLS-hard	78.18	78.16	77.66	77.32	76.70	76.13	75.47	74.63	73.96	72.96	72.29
RLLS-soft	78.18	78.16	**77.66**	**77.32**	**76.70**	**76.13**	**75.47**	**74.63**	73.96	72.96	72.29
MAPLS-APL-hard (**ours**)	78.83	78.31	77.30	76.66	75.86	75.37	74.75	74.26	74.13	74.05	74.02
MAPLS-APL-soft (**ours**)	**79.05**	**78.46**	77.48	76.81	76.06	75.56	74.99	74.58	**74.48**	**74.32**	**74.42**

Table 24. Top1 Accuracy on ImageNet dataset, with Ordered Long-Tailed test set with ordered Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 20 different shuffled and random sampled test set.

Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	2.62	3.72	6.27	8.87	14.92	23.47	37.96	67.16	99.08	153.10	201.93
MLLS-soft	51.76	48.34	54.89	45.71	84.10	77.53	107.90	137.84	184.96	251.67	262.78
BBSE-hard	$9.4e^6$	$1.0e^7$	$2.0e^7$	$5.7e^7$	$5.4e^9$	$9.1e^5$	$1.8e^6$	$1.4e^7$	$7.0e^4$	$1.7e^8$	$1.9e^8$
BBSE-soft	1.67	2.01	2.69	2.80	5.54	10.55	19.80	42.96	63.91	102.84	152.85
RLLS-hard	0.08	0.11	0.87	2.77	9.14	18.95	35.57	71.96	112.43	184.69	252.85
RLLS-soft	**0.08**	**0.11**	**0.87**	2.77	9.14	18.95	35.57	71.95	112.43	184.69	253.38
MAPLS-APL-hard (**ours**)	4.34	4.22	3.55	**2.41**	**1.01**	**1.70**	6.53	23.03	44.85	87.83	130.65
MAPLS-APL-soft (**ours**)	14.28	14.02	12.89	10.49	6.59	3.65	**4.21**	**12.34**	**26.16**	**54.83**	**83.57**

Table 25. Performance of $(w - \hat{w})^2/K$ on ImageNet-LT dataset, with ordered Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 20 different shuffled and random sampled test set.

Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	**65.24**	**62.55**	**58.48**	54.97	49.59	45.31	40.94	35.22	31.31	26.56	23.58
MLLS-hard	61.78	59.10	55.42	52.70	48.93	46.47	44.04	41.27	39.82	38.30	37.39
MLLS-soft	60.86	58.45	54.70	52.13	48.56	46.34	44.11	41.66	40.41	39.30	39.03
BBSE-hard	34.03	33.20	25.93	24.53	19.03	26.15	23.99	16.85	28.03	15.67	12.92
BBSE-soft	63.45	60.95	57.47	54.86	51.03	48.23	45.67	42.47	40.42	37.94	36.34
RLLS-hard	65.24	62.55	58.48	54.97	49.59	45.31	40.94	35.22	31.31	26.56	23.58
RLLS-soft	65.24	62.55	58.48	54.97	49.59	45.31	40.94	35.22	31.31	26.56	23.58
MAPLS-APL-hard (**ours**)	62.92	60.67	57.72	55.56	52.51	50.31	48.05	45.09	43.33	41.31	39.79
MAPLS-APL-soft (**ours**)	62.47	60.34	57.58	**55.44**	**52.50**	**50.32**	**48.33**	**45.69**	**44.21**	**42.50**	**41.57**

Table 26. Top1 Accuracy on ImageNet-LT dataset, with Ordered Long-Tailed test set with ordered Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 20 different shuffled and random sampled test set.

G.2. Shuffled Long-Tailed test set

Dataset	ImageNet					ImageNet-LT				
Test imbalance ratio	50	20	10	5	2	50	20	10	5	2
MLLS-hard	0.1210	0.1102	0.1001	0.0868	0.0766	36.09	34.49	30.57	26.90	24.42
MLLS-soft	**0.1121**	0.0972	0.0868	0.0721	0.0637	80.66	82.10	84.92	81.54	76.59
BBSE-hard	0.1285	0.1020	0.0871	0.0699	0.0581	$3.2e^5$	$1.8e^6$	$1.4e^6$	$2.0e^7$	$4.8e^5$
BBSE-soft	0.1305	0.1086	0.0969	0.0790	0.0671	28.00	25.48	18.04	15.86	12.07
RLLS-hard	1.1450	0.7160	0.4436	0.2244	0.0473	45.00	38.77	29.99	24.18	19.96
RLLS-soft	1.1450	0.7160	0.4436	0.2244	0.0473	45.00	38.77	29.98	24.18	19.96
MAPLS-APL-hard (**Ours**)	0.1236	0.1006	0.0816	0.0633	0.0482	20.25	16.62	**10.26**	**6.18**	**2.62**
MAPLS-APL-soft (**Ours**)	0.1144	**0.0904**	**0.0710**	**0.0521**	**0.0370**	**19.48**	**16.39**	11.23	7.58	4.43

Table 27. Performance of label shift estimation error $(w - \hat{w})^2/K$ on ImageNet and ImageNet-LT dataset, with shuffled Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 50 different shuffled and random sampled test set.

Dataset	ImageNet					ImageNet-LT				
Test imbalance ratio	50	20	10	5	2	50	20	10	5	2
Baseline	76.02	76.05	76.10	**76.17**	**76.10**	45.40	45.24	45.16	45.30	45.31
MLLS-hard	78.12	76.98	76.21	75.63	75.04	51.89	50.05	48.28	47.42	46.56
MLLS-soft	78.33	77.21	76.46	75.90	75.28	52.05	50.02	48.34	47.23	46.53
BBSE-hard	77.91	77.03	76.40	75.91	75.36	33.27	29.10	26.63	17.97	31.62
BBSE-soft	77.74	76.82	76.26	75.81	75.22	52.40	51.05	49.86	49.01	48.37
RLLS-hard	76.02	76.05	76.10	76.17	76.10	45.40	45.24	45.16	45.30	45.31
RLLS-soft	76.02	76.05	76.10	76.17	76.10	45.40	45.24	45.16	45.30	45.31
MAPLS-APL-hard (**Ours**)	78.21	77.16	76.46	75.93	75.36	53.36	52.10	51.09	50.67	50.30
MAPLS-APL-soft (**Ours**)	**78.38**	**77.37**	**76.69**	76.14	75.59	**53.78**	**52.35**	**51.46**	**50.85**	**50.47**

Table 28. Top1 Accuracy Performance on ImageNet and ImageNet-LT dataset, with shuffled Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 50 different shuffled and random sampled test set.

G.3. Dirichlet Shifted test set

Dataset	ImageNet						ImageNet-LT					
α	10.0			1.0			10.0			1.0		
Test Sample No.	12500	25000	37500	12500	25000	37500	12500	25000	37500	12500	25000	37500
MLLS-hard	0.1111	0.0848	0.0778	0.1299	0.1113	0.1071	28.44	26.03	24.66	38.21	36.18	35.89
MLLS-soft	0.0981	0.0721	0.0643	**0.1154**	**0.0977**	**0.0920**	91.28	83.62	92.95	82.62	84.23	84.28
BBSE-hard	0.0869	0.0661	0.0608	0.1285	0.1173	0.1118	$4.8e^5$	$1.0e^7$	$5.7e^6$	$1.7e^6$	$1.2e^10$	$1.5e^7$
BBSE-soft	0.1052	0.0769	0.0689	0.1366	0.1177	0.1099	13.84	12.89	13.15	28.30	27.75	27.42
RLLS-hard	0.1159	0.1122	0.1089	1.1020	1.0607	1.0330	21.98	21.05	21.21	46.05	45.75	45.50
RLLS-soft	0.1159	0.1122	0.1089	1.1020	1.0607	1.0330	21.98	21.05	21.21	46.05	45.75	45.50
MAPLS-APL-hard (**Ours**)	0.0736	0.0570	0.0524	0.1283	0.1142	0.1123	**4.72**	**3.86**	**3.72**	21.16	20.68	20.50
MAPLS-APL-soft (**Ours**)	**0.0628**	**0.0465**	**0.0413**	0.1160	0.1025	0.0999	6.62	5.66	5.64	**18.94**	**18.75**	**19.32**

Table 29. Performance of label shift estimation error $(w - \hat{w})^2/K$ on ImageNet and ImageNet-LT dataset, with Dirichlet shifted test set with $\alpha = 1.0, 10.0$ and sample number $12500, 25000, 37500$. Each reported values are averaging over 50 different shuffled and random sampled test set.

Dataset	ImageNet						ImageNet-LT					
α	10.0			1.0			10.0			1.0		
Test Sample No.	12500	25000	37500	12500	25000	37500	12500	25000	37500	12500	25000	37500
Baseline	**76.13**	**76.15**	**76.11**	76.21	76.20	75.98	45.26	45.34	45.28	45.27	45.26	45.16
MLLS-hard	74.80	75.20	75.28	78.01	78.08	77.93	45.94	46.68	46.95	50.91	51.20	51.13
MLLS-soft	75.09	75.44	75.53	78.25	**78.37**	**78.16**	45.70	46.58	46.83	51.04	51.23	51.27
BBSE-hard	75.22	75.51	75.58	77.94	78.02	77.83	31.15	20.76	24.76	30.57	18.31	25.88
BBSE-soft	75.00	75.36	75.47	77.88	77.98	77.75	48.13	48.51	48.66	51.65	51.87	51.70
RLLS-hard	76.13	76.15	76.11	76.21	76.20	75.98	45.26	45.34	45.28	45.27	45.26	45.16
RLLS-soft	76.13	76.15	76.11	76.21	76.20	75.98	45.26	45.34	45.28	45.27	45.26	45.16
MAPLS-APL-hard (**Ours**)	75.24	75.53	75.58	78.06	78.11	77.93	49.97	50.35	50.45	52.66	52.81	52.71
MAPLS-APL-soft (**Ours**)	75.49	75.74	75.80	**78.27**	78.35	78.14	**50.13**	**50.53**	**50.59**	**53.01**	**53.12**	**53.05**

Table 30. Top1 Accuracy Performance on ImageNet and ImageNet-LT dataset, with Dirichlet shifted test set with $\alpha = 1.0, 10.0$ and sample number $12500, 25000, 37500$. Each reported values are averaging over 50 different shuffled and random sampled test set.

H. Places and Places-LT dataset results

H.1. Ordered Long-Tailed test set

Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	**0.1374**	**0.1265**	0.1218	0.1181	0.1082	0.1048	0.1160	0.1342	0.1482	0.1628	0.1855
MLLS-soft	0.1613	0.1440	0.1397	0.1288	0.1137	0.1080	0.1161	0.1324	0.1450	0.1636	0.1845
BBSE-hard	1.4165	1.5820	1.5284	1.5041	1.4975	1.4484	1.4438	1.5558	1.4226	1.5014	1.3340
BBSE-soft	0.2584	0.2204	0.2132	0.1919	0.1675	0.1576	0.1642	0.1846	0.1996	0.2221	0.2484
RLLS-hard	1.0530	0.7509	0.4046	0.2011	0.0357	0.0053	0.0654	0.2873	0.5566	1.0297	1.4574
RLLS-soft	1.0530	0.7509	0.4046	0.2011	**0.0357**	**0.0053**	0.0654	0.2873	0.5566	1.0297	1.4574
MAPLS-APL-hard (**ours**)	0.1642	0.1334	**0.0918**	**0.0642**	0.0398	0.0331	**0.0425**	**0.0707**	**0.1002**	**0.1423**	**0.1836**
MAPLS-APL-soft (**ours**)	0.1743	0.1426	0.1060	0.0795	0.0529	0.0445	0.0553	0.0843	0.1148	0.1582	0.2019

Table 31. Performance of $(w - \hat{w})^2/K$ on Places dataset, with Ordered Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 20 different shuffled and random sampled test set.

Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	55.35	55.84	56.08	56.33	56.72	56.77	56.85	56.93	56.96	56.90	56.89
MLLS-hard	**60.70**	59.45	57.50	56.27	55.66	55.53	55.87	57.11	58.53	60.39	61.94
MLLS-soft	60.64	**59.47**	57.60	56.55	55.77	55.59	55.92	57.21	58.53	**60.50**	**62.13**
BBSE-hard	55.99	54.55	52.72	51.75	50.92	50.82	51.21	52.75	54.14	56.20	58.05
BBSE-soft	60.08	58.80	57.00	55.90	55.26	55.06	55.29	56.52	57.87	59.94	61.69
RLLS-hard	55.35	55.84	56.08	56.33	56.72	56.77	56.85	56.93	56.96	56.90	56.89
RLLS-soft	55.35	55.84	56.08	56.33	**56.72**	**56.77**	**56.85**	56.93	56.96	56.90	56.89
MAPLS-APL-hard (**ours**)	60.38	59.35	57.77	56.84	56.32	56.31	56.61	57.61	**58.84**	60.44	61.71
MAPLS-APL-soft (**ours**)	60.34	59.37	**57.84**	**56.91**	56.35	56.24	56.55	57.63	58.75	60.41	61.78

Table 32. Top1 Accuracy on Places dataset, with Ordered Long-Tailed test set with ordered Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 20 different shuffled and random sampled test set.

Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
MLLS-hard	3.4675	5.4574	9.5617	17.7925	39.9483	70.8934	120.5449	223.9239	333.2823	525.0909	698.2303
MLLS-soft	7.4563	8.1166	13.6317	27.9751	47.4603	104.0224	158.3961	291.3247	373.9417	628.1687	818.2384
BBSE-hard	$1.9e^5$	$4.5e^5$	$7.6e^5$	$1.9e^5$	$1.0e^5$	$9.7e^3$	$6.3e^4$	$8.6e^4$	$8.0e^6$	$2.3e^5$	$1.4e^5$
BBSE-soft	1.1017	2.2796	5.3919	9.9850	26.8099	51.1293	85.8800	171.5911	255.1838	409.3291	571.0979
RLLS-hard	**0.2672**	**1.0802**	**4.8371**	12.2802	34.8413	68.2377	122.7763	237.4952	361.1785	561.5883	769.9876
RLLS-soft	0.2672	1.0803	4.8413	12.2805	34.8417	68.2382	122.7766	237.4960	361.1791	575.2623	769.9874
MAPLS-APL-hard (**ours**)	10.4254	9.2915	6.1466	**3.0678**	**1.2933**	6.9923	26.4801	85.6515	162.4542	311.1073	455.9404
MAPLS-APL-soft (**ours**)	23.6750	22.3458	18.6892	14.4898	5.7818	**3.3824**	**10.1801**	**48.5159**	**97.1170**	**218.3957**	**358.1589**

Table 33. Performance of $(w - \hat{w})^2/K$ on Places-LT dataset, with ordered Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 20 different shuffled and random sampled test set.

Order	Forward					Uniform	Backward				
Imbalance Ratio	50	25	10	5	2	1	2	5	10	25	50
Baseline	**43.36**	41.25	38.04	35.11	31.06	27.92	24.76	20.89	18.26	15.49	13.46
MLLS-hard	42.26	40.46	37.78	35.67	32.95	31.08	29.22	26.85	25.30	23.70	22.57
MLLS-soft	41.77	39.90	37.20	35.06	32.43	30.53	28.72	26.58	25.50	24.11	22.85
BBSE-hard	28.42	28.65	28.39	27.83	26.37	26.79	24.51	23.09	16.69	17.60	18.44
BBSE-soft	43.00	41.12	38.32	36.18	33.16	30.94	28.75	26.16	24.34	22.31	20.54
RLLS-hard	43.36	41.25	38.04	35.11	31.06	27.92	24.76	20.89	18.26	15.72	13.46
RLLS-soft	43.36	41.25	38.04	35.11	31.06	27.92	24.76	20.89	18.26	15.49	13.46
MAPLS-APL-hard (**ours**)	42.68	**41.34**	**39.55**	**38.01**	**36.04**	34.48	32.80	30.49	28.68	26.63	25.06
MAPLS-APL-soft (**ours**)	42.48	41.15	39.32	37.78	36.04	**34.58**	**32.97**	**30.87**	**29.35**	**27.41**	**25.63**

Table 34. Top1 Accuracy on Places-LT dataset, with Ordered Long-Tailed test set with ordered Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 20 different shuffled and random sampled test set.

H.2. Shuffled Long-Tailed test set

Dataset	Places					Places-LT				
Test imbalance ratio	50	20	10	5	2	50	20	10	5	2
MLLS-hard	**0.1651**	0.1566	0.1252	0.1153	0.1105	115.1280	104.1143	98.8763	82.3951	74.4152
MLLS-soft	0.1727	0.1622	0.1348	0.1299	0.1182	142.2334	141.6962	104.8241	107.7355	101.4559
BBSE-hard	1.7527	1.8203	1.3491	1.5414	1.6365	$6.8e^6$	$4.7e^5$	$2.6e^4$	$2.3e^6$	$1.1e^6$
BBSE-soft	0.2757	0.2466	0.2011	0.1935	0.1693	102.0397	88.2631	81.8726	64.8671	55.9274
RLLS-hard	1.0996	0.6918	0.4281	0.2214	0.0470	140.2524	118.2621	107.5960	84.7745	73.4672
RLLS-soft	1.0996	0.6918	0.4281	0.2214	0.0470	140.0257	118.2626	107.5964	84.7749	73.4666
MAPLS-APL-hard (**Ours**)	0.1730	**0.1304**	**0.0877**	**0.0613**	**0.0412**	72.4649	52.5443	39.3145	21.9486	10.3899
MAPLS-APL-soft (**Ours**)	0.1891	0.1437	0.1058	0.0794	0.0558	**65.7596**	**46.7970**	**31.5570**	**18.5238**	**6.2758**

Table 35. Performance of label shift estimation error $(w - \hat{w})^2/K$ on Places and Places-LT dataset, with shuffled Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 50 different shuffled and random sampled test set.

Dataset	Places					Places-LT				
Test imbalance ratio	50	20	10	5	2	50	20	10	5	2
Baseline	56.36	56.31	57.00	56.95	**56.78**	27.81	27.80	27.45	28.03	27.80
MLLS-hard	61.40	59.15	58.28	57.07	55.71	35.79	34.48	32.32	32.13	31.21
MLLS-soft	**61.50**	59.27	58.40	57.02	55.77	35.39	34.01	31.83	31.41	30.65
BBSE-hard	56.43	54.25	53.63	52.27	50.89	19.31	22.21	26.80	20.89	20.47
BBSE-soft	60.68	58.42	57.69	56.31	55.23	34.42	33.43	31.69	31.79	30.94
RLLS-hard	56.36	56.31	57.00	56.95	56.78	27.81	27.80	27.45	28.03	27.80
RLLS-soft	56.36	56.31	57.00	56.95	56.78	27.81	27.80	27.45	28.03	27.80
MAPLS-APL-hard (**Ours**)	61.23	**59.31**	**58.67**	**57.59**	56.46	36.51	35.94	34.52	34.85	34.46
MAPLS-APL-soft (**Ours**)	61.19	59.28	58.58	57.46	56.40	**36.55**	**35.97**	**34.58**	**34.87**	**34.55**

Table 36. Top1 Accuracy Performance on Places and Places-LT dataset, with shuffled Long-Tailed test set that have imbalance ratio $R = \{50, 10, 5, 2\}$. Each reported values are averaging over 50 different shuffled and random sampled test set.

H.3. Dirichlet Shifted test set

Dataset	Places						Palces-LT					
α	10.0			1.0			10.0			1.0		
Test Sample No.	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500
MLLS-hard	0.3541	0.2263	0.1809	0.3312	0.2151	0.1778	82.8776	81.0168	80.1520	128.4708	123.9672	129.7458
MLLS-soft	0.3985	0.2443	0.1953	0.3694	0.2422	0.1884	104.2574	106.8853	90.7455	148.7993	146.5841	156.8282
BBSE-hard	1.7579	1.6217	1.6357	1.9250	1.8103	1.7504	$1.6e^6$	$9.8e^4$	$9.6e^9$	$1.6e^5$	$2.2e^4$	$49.9e^5$
BBSE-soft	0.5594	0.3516	0.2858	0.5881	0.3807	0.3104	59.0054	58.1415	56.6426	108.3152	109.0846	112.5832
RLLS-hard	**0.1276**	0.1158	0.1133	1.1764	1.0826	1.069	78.5794	77.8595	77.9273	150.6753	148.8841	152.7133
RLLS-soft	0.1276	0.1158	0.1133	1.1764	1.0826	1.069	78.5799	77.8603	77.9278	150.6757	148.8845	152.7137
MAPLS-APL-hard (**Ours**)	0.1707	**0.1038**	**0.0818**	**0.2716**	**0.1945**	**0.1684**	19.0960	17.1211	16.0253	79.9565	77.0303	79.3220
MAPLS-APL-soft (**Ours**)	0.1940	0.1219	0.0973	0.2934	0.2150	0.1827	**13.4071**	**12.2247**	**10.5204**	**69.5078**	**69.0457**	**69.4423**

Table 37. Performance of label shift estimation error $(w - \hat{w})^2/K$ on Places and Places-LT dataset, with Dirichlet shifted test set with $\alpha = 1.0, 10.0$ and sample number $12500, 25000, 37500$. Each reported values are averaging over 50 different shuffled and random sampled test set.

Dataset	Places						Places-LT					
α	10.0			1.0			10.0			1.0		
Test Sample No.	2500	5000	7500	2500	5000	7500	2500	5000	7500	2500	5000	7500
Baseline	56.80	56.90	56.79	56.64	56.81	56.71	27.80	27.94	27.85	28.18	27.94	28.03
MLLS-hard	53.58	54.98	55.36	59.72	60.61	60.90	30.58	30.94	31.07	34.72	35.28	35.33
MLLS-soft	53.40	55.00	55.35	59.59	60.56	**60.93**	30.19	30.33	30.64	34.35	35.04	34.98
BBSE-hard	49.90	50.81	50.85	55.52	55.93	56.22	20.73	25.53	16.33	27.49	29.74	25.94
BBSE-soft	52.47	54.05	54.47	58.65	59.71	60.10	33.80	34.08	34.09	33.80	34.08	34.09
RLLS-hard	56.80	56.90	56.79	56.64	56.81	56.71	27.80	27.94	27.85	28.18	27.94	28.03
RLLS-soft	**56.80**	**56.90**	**56.79**	56.64	56.81	56.71	27.80	27.94	27.85	28.18	27.94	28.03
MAPLS-APL-hard (**Ours**)	55.04	56.08	56.30	**59.94**	**60.67**	60.88	34.08	34.34	34.30	36.20	36.49	36.39
MAPLS-APL-soft (**Ours**)	54.96	56.08	56.25	59.88	60.61	60.85	**34.16**	**34.29**	**34.36**	**36.27**	**36.58**	**36.44**

Table 38. Top1 Accuracy Performance on Places and Places-LT dataset, with Dirichlet shifted test set with $\alpha = 1.0, 10.0$ and sample number $12500, 25000, 37500$. Each reported values are averaging over 50 different shuffled and random sampled test set.