Int. J. Appl. Math. Comput. Sci., 2025, Vol. 35, No. 3, 507–517
DOI: 10.61822/amcs-2025-0036

ACCOUNTING FOR LABEL SHIFT OF POSITIVE UNLABELED DATA UNDER SELECTION BIAS

JAN MIELNICZUK [a,b,*], ADAM WAWRZEŃCZYK [a]

[a]Institute of Computer Science
Polish Academy of Sciences
ul. Jana Kazimierza 5, 01-248 Warsaw, Poland
e-mail: jan.mielniczuk@ipipan.waw.pl

[b]Faculty of Mathematics and Information Sciences
Warsaw University of Technology
ul. Koszykowa 5, 00-662 Warsaw, Poland

We consider the scenario when two samples of positive unlabeled (PU) data are available and for the second sample the change in prior probability of classes occurs while distributions of predictors in classes remain the same (label shift setting). The selection of positive elements may be object-dependent. We study the properties of the underlying probabilistic structure under the novel augmented PU scenario, proving in particular that label shift occurs also for unlabeled populations. We introduce and investigate an estimator of prior probability for label-shifted population. Furthermore, in this case we construct and analyze behavior of Bayes classifier in this setting. It turns out to be a Bayes classifier for the unlabeled class with a modified threshold. This gives rise to its three empirical counterparts which are compared on benchmark data sets.

Keywords: positive unlabeled learning, label shift, augmented positive unlabeled data, selection bias, Bayes classifier, accuracy.

1. Label-shift augmented positive unlabeled model: An introduction

In the paper we consider two departures from a classic classification model which frequently occur in practice and need to be accounted for. The first one is that data is often only partially available, in particular information on class indicators may be restricted. More specifically, for positive unlabeled (PU) data considered here, labeling information is available only for a subset of observations belonging to the positive class. Moreover, the mechanism revealing class indicators may depend on observation's features (selection bias) and this influences properties of standard classifiers. The second departure is that the distribution of data on which classifiers are trained may differ from the distribution of data to be classified, in particular a value of prior probability may shift. Here, we propose modeling scenario which incorporates these nonstandard data features and show how to construct optimal classifiers in this case.

We first introduce basic notations. Let X be a random variable corresponding to feature vector, $Y \in \{-1, 1\}$ be a true class label and $S \in \{0, 1\}$ be an indicator of an example being labeled ($S = 1$) or not ($S = 0$). We assume that there is some unknown probability distribution P_{XYS} such that only positive examples ($Y = 1$) can be labeled, i.e., $P(S = 1|X, Y = -1) = 0$. Thus we know that $Y = 1$ when $S = 1$ but when $S = 0$, Y can be either 1 or -1. Vector X is a vector of predictors and we assume w.l.o.g. that $X \in R^p$. Inference for class indicator Y based on a sample pertaining to P_{XS} is an actively researched problem of machine learning for PU data (for a review of approaches, see, e.g., the work of Bekker and Davis (2020)).

Moreover, we consider the second vector $(\widetilde{X}, \widetilde{Y})$ such that its distribution $P_{\widetilde{X}\widetilde{Y}}$ is label-shifted distribution of P_{XY}, which means that marginal distribution of $\widetilde{Y}$ is different from that of Y, i.e.,

$$\widetilde{\pi} := P(\widetilde{Y} = 1) \neq P(Y = 1) =: \pi, \tag{1}$$

*Corresponding author

however we assume distributions in a positive and negative class (i.e., conditional distributions of predictors given class indicator) are the same for both distributions:

$$P_{\widetilde{X}|\widetilde{Y}=i} = P_{X|Y=i}, \quad i = \pm 1. \tag{2}$$

Although the proposed method is applicable also in the case $\widetilde{\pi} = \pi$, in order to streamline the presentation, we assume (1). Note that X and $\widetilde{X}$ correspond to the vectors of variables defined in the same way, which, have however different distributions in the image space, the same convention refers to Y and $\widetilde{Y}$. Since $P_X = \pi P_{X|Y=1} + (1-\pi)P_{X|Y=-1}$ and analogous expression holds for $P_{\widetilde{X}}$, marginal distributions of X and $\widetilde{X}$, in contrast to conditional distributions, also differ: $P_X \neq P_{\widetilde{X}}$.

Such situation occurs frequently in practice. Consider, e.g., the anti-causal case when X denotes symptoms of a certain disease and $Y = 1$ when it occurs. Then when prevalence of disease changes but characteristics of the symptoms of disease do not, this corresponds to label shift scenario. We assume that the second vector is not fully observable either, in the sense that, as in the case of the first vector, positive labels are only partially labeled, and denote labeling variable by $\widetilde{S}$ in this case.

Such a scenario corresponds to practical situation when, e.g., one's aim is to screen people suffering from hypertension. People who check their blood pressure regularly and if its abnormal, report this to a doctor and are treated. Remaining group consists of people who are healthy and those who have hypertension but do not contact a doctor. Having such data for two consecutive periods of time, we would like to detect those in the second group who suffer from hypertension, but allowing for a possible change of its prevalence in the population considered. Note that this scenario is applicable in many other cases like COVID-19 detection based on certain number of symptoms (coughing, difficulty of breathing) and patient's characteristics, without the necessity of performing COVID-19 test.

We consider a realistic setting when labeling mechanism may be object dependent, i.e., *selection bias* occurs. This corresponds to the general selected at random (SAR) scenario considered for PU data (see the work of Bekker and Davis (2020) for a discussion of various labeling mechanisms for PU data). In selected completely at random (SCAR), the case labeling mechanism is independent of features (see, e.g., Wawrzeńczyk and Mielniczuk, 2022). In the following, we assume that the labeling which censors positive observations acts in the same manner in the first and in the second case, namely

$$\begin{aligned} e(x) &:= P(S=1|Y=1, X=x) \\ &= P(\widetilde{S}=1|\widetilde{Y}=1, \widetilde{X}=x) =: \widetilde{e}(x), \end{aligned} \tag{3}$$

i.e., that propensity scores $e(x)$ and $\widetilde{e}(x)$ are the same and they will be denoted by $e(x)$ henceforth. Note that as P_{XYS} is characterized by marginal distribution P_Y and conditional distributions $P_{X|Y}$ and $P_{S|XY}$, the probabilistic structures of P_{XYS} and $P_{\widetilde{X}\widetilde{Y}\widetilde{S}}$ are uniquely determined by the assumptions above. We also define posterior probabilities of $Y = 1$ given $X = x$ as $y(x) = P(Y = 1|X = x)$ and $s(x) = P(S = 1|X = x)$; conditional probabilities $\widetilde{y}(x)$ and $\widetilde{s}(x)$ are defined analogously. We denote by f_X either a density of X or its probability mass function and the same convention applies to $f_{X|S=1}$.

Note that due to the definition of a conditional probability we have

$$\begin{aligned} s(x) &= P(S=1|X=x) \\ &= P(S=1|Y=1, X=x)P(Y=1|X=x) \\ &= e(x)y(x), \end{aligned} \tag{4}$$

and we analogously obtain

$$\widetilde{s}(x) = e(x)\widetilde{y}(x). \tag{5}$$

Assume that $\mathcal{D} = (Y_i, X_i, S_i), i = 1, \ldots, n$ is an iid sample drawn from distribution P_{XYS} and $\widetilde{\mathcal{D}} = (\widetilde{Y}_i, \widetilde{X}_i, \widetilde{S}_i), i = 1, \ldots, m$ is iid sample drawn from distribution $P_{\widetilde{X}\widetilde{Y}\widetilde{S}}$ independently of $\mathcal{D}$. Observed data consists of $(X_i, S_i), i = 1, \ldots, n$ and $(\widetilde{X}_i, \widetilde{S}_i), i = 1, \ldots, m$, thus in both cases only partial information on labels is available. Our aim is to construct a classification rule and predict class indicator $\widetilde{Y} = \pm 1$ for observations in $\widetilde{\mathcal{D}}$. Note that since $\widetilde{S}_i = 1 \Rightarrow \widetilde{Y}_i = 1$ the task reduces to classification of unlabeled observations ($\widetilde{S}_i = 0$) in $\widetilde{\mathcal{D}}$. We note in passing that the considered setting corresponds to single-training sample training scenario, and case-control setting is also frequently considered for PU data (Kiryo *et al*., 2017).

We emphasize that under the standard label shift probability scenario, one observes samples $(X_i, Y_i), i = 1, \ldots, n$ and $\widetilde{X}_i, i = 1, \ldots, m$ and the task is to predict class indicators for the second sample which is shifted and for which class indicators are missing. For representative examples of methods designed for such scenario we refer to the works of Saerens *et al*. (2002), Lipton *et al*. (2018) and Garg *et al*. (2020); see also the works of Iyer *et al*. (2014), Vaz *et al*. (2019), Ye *et al*. (2024) and the references therein. We note that estimation of shifted prior probability (quantification task) is of importance in business applications (see, e.g., González *et al*., 2017). The scenario considered here, in the case when no label shift occurs, i.e., $P_{XY} = P_{\widetilde{X}\widetilde{Y}}$, called augmented PU scenario, was recently introduced by Wawrzeńczyk and Mielniczuk (2024), who considered classification of new observation (X, S) following P_{XS}. To the best of our knowledge, despite its practical

importance, label shift for augmented PU data has not been analysed in the literature.

The main contributions of the paper are: (i) we construct a model for PU data which takes into account selection bias and potential label shift of target data; (ii) we establish properties of probabilistic structure of main entities in the constructed model and a form of Bayes classifier for unlabeled observations in target data; (iii) we consider empirical counterparts of the Bayes rule using different estimators of prior probability for target data and compare their behaviour on real data sets.

2. Main theoretical results

Below we present some basic facts concerning the label-shift augmented PU model (Section 2.1) and form of the Bayes classifiers under this scenario (Section 2.2).

2.1. General results. Lemma 1 below describes the basic facts on interplay between P_{XYS} and $P_{\widetilde{X}\widetilde{Y}\widetilde{S}}$, which will be useful for construction of classification rule based on $(\widetilde{X}, \widetilde{S})$. In particular, we prove in part (ii) that distribution $P_{\widetilde{X}\widetilde{Y}|\widetilde{S}=0}$ is label-shifted distribution of $P_{XY|S=0}$. Denote by $c = P(S = 1|Y = 1)$ and $\widetilde{c} = P(\widetilde{S} = 1|\widetilde{Y} = 1)$ overall conditional probabilities of being labeled for the first and second sample, respectively. Moreover, define odds of positive class occurring for the first sample as $OD(x) = P(Y = 1|x)/P(Y = -1|x)$ with $\widetilde{OD}(x)$ defined analogously for the second sample. Odds Ratio $OR(x)$ equals $OR(x) = \widetilde{OD}(x)/OD(x)$.

Lemma 1.
The following equalities hold:

(i) $\widetilde{c} = c$.

(ii) Assume that $\widetilde{\pi} \neq \pi$. *Then distribution* $P_{\widetilde{X}\widetilde{Y}|\widetilde{S}=0}$ *is label shifted distribution of* $P_{XY|S=0}$. *Namely, we have*

$$P(\widetilde{Y} = 1|\widetilde{S} = 0) = \frac{\widetilde{\pi} - c\widetilde{\pi}}{1 - c\widetilde{\pi}},$$
$$\neq P(Y = 1|S = 0) = \frac{\pi - c\pi}{1 - c\pi},$$

and

$$f_{\widetilde{X}|\widetilde{Y}=1,\widetilde{S}=0}(x) = f_{X|Y=1,S=0}(x),$$
$$f_{\widetilde{X}|\widetilde{Y}=0,\widetilde{S}=0}(x) = f_{X|Y=0,S=0}(x).$$

(iii) $f_{X|S=1}(x) = f_{\widetilde{X}|\widetilde{S}=1}(x)$.

(iv) $OD(x)\frac{1-\pi}{\pi} = \widetilde{OD}(x)\frac{1-\widetilde{\pi}}{\widetilde{\pi}} \equiv OR(x) = \frac{\widetilde{\pi}}{1-\widetilde{\pi}} \times \frac{1-\pi}{\pi}$.

Proof.

(i) Note that

$$c = \int P(S = 1|Y = 1, X = x) f_{X|Y=1}(x)\, dx$$
$$= \mathbb{E}_{X|Y=1}\, e(X) = \mathbb{E}_{\widetilde{X}|\widetilde{Y}=1}\, e(\widetilde{X}) = \widetilde{c},$$

where the first equality follows from definitions of c and $e(x)$, the second from assumed equality of distributions within classes, and the last one from equality $\widetilde{e}(x) = e(x)$.

(ii) To prove the first part, note that $P(\widetilde{S} = 0) = 1 - c\widetilde{\pi}$, which yields expressions for conditional probabilities. Moreover, inequality of the conditional probabilities follows from strict monotonicity of the function $f(a) = a/(1 - ca)$ for $a \in (0, 1)$.

For the second part, note that

$$f_{X|Y=1,S=0}(x)$$
$$= \frac{P(Y = 1, X = x)P(S = 0|Y = 1, X = x)}{P(Y = 1, S = 0)}$$
$$= \frac{f_{X|Y=1}(x)\pi(1 - e(x))}{P(Y = 1, S = 0)} = \frac{f_{X|Y=1}(x)\pi(1 - e(x))}{\pi - \pi c}$$
$$= f_{X|Y=1}(x)(1 - e(x))(1 - c)^{-1}.$$

As analogous formula holds for $f_{\widetilde{X}|\widetilde{Y}=1,\widetilde{S}=0}(x)$, in view of (3), $f_{X|Y=1} = f_{\widetilde{X}|\widetilde{Y}=1}$ and $c = \widetilde{c}$, the first part is proved. The second is even more straightforward as

$$f_{X|Y=0,S=0}(x) = f_{X|Y=0}(x)$$
$$= f_{\widetilde{X}|\widetilde{Y}=0}(x) = f_{\widetilde{X}|\widetilde{Y}=0,\widetilde{S}=0}(x).$$

(iii) Note that in view of (4) we have

$$f_{X|S=1}(x) = \frac{s(x)f_X(x)}{P(S = 1)} = \frac{y(x)e(x)f_X(x)}{P(S = 1)}$$
$$= \frac{P(Y = 1, X = x)}{f_X(x)} \times \frac{e(x)f_X(x)}{P(S = 1)}$$
$$= f_{X|Y=1}(x)\frac{\pi e(x)}{P(S = 1)}.$$

Replacing in the last expression $f_{X|Y=1}(x)$ by $f_{\widetilde{X}|\widetilde{Y}=1}(x)$ and repeating the above line of argument backwards we obtain

$$f_{X|S=1}(x) = f_{\widetilde{X}|\widetilde{S}=1}(x)\frac{\pi}{P(S = 1)}\frac{P(\widetilde{S} = 1)}{\widetilde{\pi}}$$
$$= f_{\widetilde{X}|\widetilde{S}=1}(x) \times \widetilde{c}/c = f_{\widetilde{X}|\widetilde{S}=1}(x),$$

where the last equality follows from (i).

(iv) Reasoning as before, we note that

$$\widetilde{y}(x) = \frac{f_{\widetilde{X}|\widetilde{Y}=1}(x)\widetilde{\pi}}{f_{\widetilde{X}}(x)} = \frac{y(x)f_X(x)\widetilde{\pi}/\pi}{f_{\widetilde{X}}(x)},$$

and

$$1-\widetilde{y}(x)=\frac{(1-y(x))f_X(x)(1-\widetilde{\pi})/(1-\pi)}{f_{\widetilde{X}}(x)}.$$

Dividing the expression above yields the conclusion.

■

Remark 1.

(i) We note that important equality in Lemma 1 (iii) can be intuitively justified by noting that labeled $(S=1)$ observation $X=x$ is picked from the strata $Y=1$ described by distribution $f_{X|Y=1}$ with probability $e(x)$. In the case of label-shifted distribution, distributions of positive class and labeling mechanism are the same as for P_{XYS}.

(ii) Note that Lemma 1 (i) implies, as $c=P(S=1)/P(Y=1)$, that if $\widetilde{\pi}>\pi$ then $P(\widetilde{S}=1)>P(S=1)$ and vice versa. Moreover, proportion of positives to negatives equals $\pi(1-c)/(1-\pi)$ for unlabeled population $S=0$ and $\widetilde{\pi}(1-c)/(1-\widetilde{\pi})$ for $\widetilde{S}=0$ (see Lemma 1 (ii)).

We also note that the stronger property than Lemma 1.(i) holds, namely $P(S=1|Y=1,X\in A)=P(\widetilde{S}=1|\widetilde{Y}=1,\widetilde{X}\in A)$. Moreover, note that no label-shift situation $\widetilde{\pi}=\pi$ is equivalent in the view of Lemma 1 (i) to $P(S=1)=P(\widetilde{S}=1)$ which can be routinely tested using difference of two binomial proportion test (this does not require knowledge of prior π).

In the following, we assume that the prior probability $\pi=P(Y=1)$ is known. This is reasonable assumption when $Y=1$ corresponds to disease and its prevalence can be estimated with arbitrary accuracy. We note in passing that in this case distribution of negative examples $P_{X|Y=0}=(1-\pi)^{-1}\left(P_X-\pi P_{X|Y=1}\right)$ is identifiable although no sample pertaining to it is available. Denote $\gamma=P(S=1)$ and $\widetilde{\gamma}=P(\widetilde{S}=1)$. Then Lemma 1 (i) can be rewritten as

$$\widetilde{\pi}=\frac{P(\widetilde{S}=1)}{P(S=1)}\times\pi, \tag{6}$$

thus yielding plug-in estimator of $\widetilde{\pi}$:

$$\begin{aligned}\widehat{\widetilde{\pi}}&=\frac{\widehat{\widetilde{\gamma}}}{\widehat{\gamma}}\mathbb{I}\{\widehat{\gamma}>0\}\times\pi\\&=\frac{\#(i:\,\widetilde{S}_i=1)/m}{\#(i:\,S_i=1)/n}\mathbb{I}\{\widehat{\gamma}>0\}\times\pi.\end{aligned} \tag{7}$$

Lemma 2 below lists the basic properties of $\widehat{\widetilde{\pi}}$.

Lemma 2.

(i) We have for any $\delta>0$ that with probability at least $1-\delta$

$$|\widehat{\widetilde{\pi}}-\widetilde{\pi}|\le\frac{1}{c}\left(\frac{1}{\widehat{\widetilde{\gamma}}}\sqrt{\frac{1}{n}\log\left(\frac{4}{\delta}\right)}+\sqrt{\frac{1}{m}\log\left(\frac{4}{\delta}\right)}\right) \tag{8}$$

and the rate of almost sure convergence of $\widehat{\widetilde{\pi}}$ to $\widetilde{\pi}$ is $\min(n,m)^{-1/2}$.

(ii) We have, with $\widetilde{\gamma}=P(\widetilde{S}=1)$

$$\mathbb{E}\,\widehat{\widetilde{\pi}}=\widetilde{\gamma}\,\mathbb{E}\left(\widehat{\gamma}^{-1}\mathbb{I}\{\widehat{\gamma}>0\}\right)\times\pi=\widetilde{\pi}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right).$$

Proof.

(i) Using $|\widehat{\widetilde{\gamma}}|\le 1$ and the triangle inequality, the proof follows routinely from

$$\begin{aligned}|\widehat{\widetilde{\pi}}-\widetilde{\pi}|&\le\pi\left\{\left|\frac{\widehat{\widetilde{\gamma}}}{\widehat{\gamma}}-\frac{\widetilde{\gamma}}{\gamma}\right|+\left|\widehat{\widetilde{\gamma}}-\widetilde{\gamma}\right|\frac{1}{\gamma}\right\}\\&\le\pi\left\{\left|\frac{\widehat{\gamma}-\gamma}{\widehat{\gamma}\gamma}\right|+\left|\widehat{\widetilde{\gamma}}-\widetilde{\gamma}\right|\frac{1}{\gamma}\right\}\\&=\frac{\pi}{\gamma}\left\{\left|\frac{\widehat{\gamma}-\gamma}{\widehat{\gamma}}\right|+\left|\widehat{\widetilde{\gamma}}-\widetilde{\gamma}\right|\right\},\end{aligned}$$

on the set $\{\widehat{\gamma}>0\}$, equality $\pi/\gamma=c^{-1}$ and application of Hoeffiding's exponential inequality applied to binomial proportions (see Proposition 2.5 of Wainwright (2019)).

(ii) Let $N_1=\#\{S_i=1\}$ and $M_1=\#\{\widetilde{S}_i=1\}$ be the sizes of labeled samples in $\mathcal{D}$ and $\widetilde{\mathcal{D}}$ respectively, and note that using independence of M_1 and N_1:

$$\begin{aligned}\mathbb{E}\,\widehat{\widetilde{\pi}}&=\mathbb{E}_{M_1}\left(\mathbb{E}\left(\widehat{\widetilde{\pi}}\middle|M_1\right)\right)\times\pi\\&=\mathbb{E}\left(\frac{M_1}{m}\frac{1}{\widehat{\widetilde{\gamma}}}\mathbb{I}\{\widehat{\gamma}>0\}\right)\times\pi\\&=\mathbb{E}\left(\frac{M_1}{m}\right)\mathbb{E}\left(\frac{1}{\widehat{\widetilde{\gamma}}}\mathbb{I}\{\widehat{\gamma}>0\}\right)\times\pi.\end{aligned}$$

Lemma 2 (ii) thus follows from Lemma 1 of Mielniczuk (1985), which implies that

$$\mathbb{E}\left(\frac{1}{\widehat{\widetilde{\gamma}}}\mathbb{I}\{\widehat{\gamma}>0\}\right)=\frac{1}{\gamma}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right),$$

and thus

$$\begin{aligned}&\mathbb{E}\left(\frac{M_1}{m}\frac{1}{\widehat{\widetilde{\gamma}}}\mathbb{I}\{\widehat{\gamma}>0\}\right)\times\pi\\&=\frac{\widetilde{\gamma}}{\gamma}\times\pi\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)=\widetilde{\pi}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right).\end{aligned}$$

■

2.2. Bayes classification rules for label-shift augmented PU data. Let $\eta(x) = P(Y = 1|S = 0, X = x)$ and $\widetilde{\eta}(x) = P(\widetilde{Y} = 1|\widetilde{S} = 0, \widetilde{X} = x)$. In the work of Wawrzeńczyk and Mielniczuk (2024), Bayes classification function $d_B^{PU}(x) = \eta(x)/(1-\eta(x))$ based on (X, S) was considered on strata $S = 0$. It follows that the corresponding Bayes rule has the following form: the observation is classified to the positive class, if the condition

$$d_B^{PU}(x) = \frac{\eta(x)}{1-\eta(x)} = \frac{y(x) - s(x)}{1 - y(x)} > 1 \qquad (9)$$

is satisfied, and to the negative class in the opposite case. Here we show that in the label shift case the rule is modified by changing the threshold of the $d_B^{PU}(x)$.

Theorem 1.

(i) The Bayes rule for $(\widetilde{X}, \widetilde{S}) = (x, 0)$ has the following form. Classify $(x, 0)$ to class $Y = 1$ if the condition

$$\begin{aligned} d_B^{PU}(x) &> \frac{\pi}{1-\pi}\frac{1-\widetilde{\pi}}{\widetilde{\pi}} \\ &= \frac{\frac{P(S=1)}{P(\widetilde{S}=1)} - \pi}{1-\pi} = \frac{\frac{\pi}{\widetilde{\pi}} - \pi}{1-\pi} =: \theta \end{aligned} \qquad (10)$$

is satisfied, or formulating the rule equivalently, $y(x) > (\theta + s(x))/(1+\theta)$.

(ii) The Bayes risk of $\widetilde{d}_B^{PU}$ equals

$$\begin{aligned} &P(\widetilde{S} = 0)\, \mathbb{E}_{\widetilde{X}|\widetilde{S}=0} \min\left(\widetilde{\eta}(\widetilde{X}), 1 - \widetilde{\eta}(\widetilde{X})\right) \\ &= \frac{1}{2}P(\widetilde{S} = 0) - \frac{1}{2}\,\mathbb{E}_{\widetilde{X}\widetilde{S}}\left|2\widetilde{\eta}(\widetilde{X}) - 1\right|. \end{aligned}$$

Note that, if no label shift occurs, then the threshold $\theta = 1$, thus the result generalizes Theorem 1 (ii) of Wawrzeńczyk and Mielniczuk (2024).

Proof.

(i) Let $\widetilde{d}_B^{PU}(x) = \widetilde{\eta}(x)/(1 - \widetilde{\eta}(x))$ be a Bayes classification function corresponding to $\widetilde{\eta}(x)$. The pertaining Bayes rule has the following form

$$\begin{aligned} 1 < \frac{\widetilde{\eta}(x)}{1-\widetilde{\eta}(x)} &= \frac{P(\widetilde{Y} = 1|\widetilde{S} = 0, \widetilde{X} = x)}{P(\widetilde{Y} = 0|\widetilde{S} = 0, \widetilde{X} = x)} \\ &= \frac{P(\widetilde{Y} = 1, \widetilde{S} = 0, \widetilde{X} = x)}{P(\widetilde{Y} = 0, \widetilde{S} = 0, \widetilde{X} = x)} \\ &= \frac{f_{\widetilde{X}}(x)}{f_{\widetilde{X}}(x)} \times \frac{\widetilde{y}(x) - \widetilde{s}(x)}{1 - \widetilde{y}(x)} = \frac{\widetilde{y}(x)(1 - e(x))}{1 - \widetilde{y}(x)} \\ &= \frac{y(x)(1-e(x))}{1-y(x)}\frac{1-\pi}{\pi}\frac{\widetilde{\pi}}{1-\widetilde{\pi}}, \end{aligned}$$

where the last equality follows from Lemma 1 (iv). This is, using Lemma 1 (i), equivalent to the following event

$$\begin{aligned} \frac{y(x) - s(x)}{1 - y(x)} &> \frac{\pi}{\widetilde{\pi}}\frac{1-\widetilde{\pi}}{1-\pi} \\ &= \frac{P(S=1)}{P(\widetilde{S}=1)}\frac{(1 - \pi P(\widetilde{S}=1)/P(S=1))}{1-\pi} \\ &= \frac{P(S=1) - \pi P(\widetilde{S}=1)}{P(\widetilde{S}=1)(1-\pi)} = \frac{\frac{P(S=1)}{P(\widetilde{S}=1)} - \pi}{1-\pi}. \end{aligned}$$

(ii) The proof follows from Theorem 1 of Wawrzeńczyk and Mielniczuk (2024).

■

Theorem 1 (i) can be explained in the following way. As $P_{\widetilde{X}|\widetilde{S}=0}$ is label shifted distribution of $P_{X|S=0}$ and decisions on $S = 1$ and and $\widetilde{S} = 1$ are error-free, the Bayes rule for the target population is Bayes rule for the source with changed threshold. The Bayes rule in question is $d_B^{PU}(x)$, and the modified threshold for label shifted population is in this case (see, e.g., Elkan (2001))

$$\frac{P(Y=1|S=0)}{P(Y=0|S=0)}\frac{P(\widetilde{Y}=0|\widetilde{S}=0)}{P(\widetilde{Y}=1|\widetilde{S}=0)} = \frac{\pi}{1-\pi}\frac{1-\widetilde{\pi}}{\widetilde{\pi}},$$

where the last equality follows from Lemma 1 (i). This coincides with (10). Note that, surprisingly at the first sight, the threshold does not depend on c. This is due to the fact that the imbalance ratio for the target and training sample equal $(\widetilde{\pi}/(1-\widetilde{\pi}))/(\pi/(1-\pi))$ is the same as imbalance ratio for their corresponding unlabeled subsamples (see Remark 1 (ii)).

Note that if $\widetilde{\pi} > \pi$ then the rule becomes less conservative than in the no-label-shift case. Moreover, it follows from the proof above that the Bayes rule which (erroneously) does not take into account the label shift will classify to positive class if

$$\frac{y(x) - s(x)}{1 - y(x)} > 1 \quad \equiv \quad \frac{\widetilde{y}(x) - \widetilde{s}(x)}{1 - \widetilde{y}(x)} > \frac{\widetilde{\pi}}{1-\widetilde{\pi}}\frac{1-\pi}{\pi}.$$

Rule in (10) yields its empirical analogue for label-shift case: classify as positive ($Y = 1$) when

$$\frac{\widehat{y}(x) - \widehat{s}(x)}{1 - \widehat{y}(x)} > \frac{\frac{N_1}{n}\frac{m}{M_1} - \pi}{1-\pi}, \qquad (11)$$

where N_1, M_1 are sizes of labeled samples in $\mathcal{D}$ and $\widetilde{\mathcal{D}}$, respectively and $\widehat{y}(x)$, $\widehat{s}(x)$ are estimators of $y(x)$ and $s(x)$ discussed below.

3. Experiments

3.1. Datasets. To estimate the performance of the label shift methods, we based our experiments on several

Table 1. Dataset statistics.

Name	Samples	Features	Class prior $\boldsymbol{\pi}$
MNIST 3v5	13454	784	0.53
MNIST OvE	70000	784	0.51
CIFAR CT	12000	512	0.50
CIFAR MA	60000	512	0.40
STL MA	13000	512	0.40

datasets with varying characteristics (summarized in Table 1). MNIST[1] dataset is considered in two variants: 3v5, when *3*'s are positive and *5*'s are negative, and OvE, where the classes are split between Odd and Even examples. Similarly, CIFAR[2] dataset is split into two tasks: Car-Truck (CT), where *automobile* (car) images are positive, and *truck*s are negative; as well as Vehicle-Animal (VA) split differentiating between vehicles (*airplane, automobile, ship and truck*) and animals (*bird, cat, deer, dog, frog and horse*). STL[3] dataset has identical classes as CIFAR, and there, only Machine-Animal (MA) split is considered. Note that (see Table 1) classes for the first four data sets are almost exactly balanced, for the remaining two the ratio of positive to negative elements is 2 to 3.

We used several settings for the label frequency: $c \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. For MNIST datasets, PU labeling was biased according to example's *boldness* score, i.e., mean value of a pixel over the whole image. The portion of the most bold positive examples which corresponded to the desired label frequency was labeled, yielding control of c. In the case of CIFAR dataset, the labeling process utilized *redness* score (defined as $r(x) = (R(x)-G(x))+(R(x) - B(x))$, where $R(\cdot), G(\cdot), B(\cdot)$ correspond to mean R, G and B channel pixel values of input image x) as a labeling measure. The same metric was used also for STL labeling.

We simulate label shift phenomenon as follows. We assume several $\widetilde{\pi}$ values ($\widetilde{\pi} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, as well as the original dataset, corresponding to $\widetilde{\pi} = \pi$). At test time we obtain label shifted dataset:

- $\widetilde{\boldsymbol{\pi}} > \boldsymbol{\pi}$. In order to increase the prior in the test dataset, we drop a portion of negative examples. We randomly sample $\frac{1-\widetilde{\pi}}{\widetilde{\pi}} \times \#(Y_{test} = 1) < \#(Y_{test} = -1)$ negative examples. As the ratio of positives to negatives is now approximately $n\pi/[n\pi(1 - \widetilde{\pi})/\widetilde{\pi})]$, we obtain a label shifted example with prior $\widetilde{\pi}$.

- $\widetilde{\boldsymbol{\pi}} < \boldsymbol{\pi}$. Symmetric case. We decrease the proportion of positive examples to $\widetilde{\pi}$ by sampling $\frac{\widetilde{\pi}}{1-\widetilde{\pi}} \times \#(Y_{test} = -1) < \#(Y_{test} = 1)$ positives.

- $\widetilde{\boldsymbol{\pi}} = \boldsymbol{\pi}$. We preserve the original test dataset.

[1] http://yann.lecun.com/exdb/mnist/.
[2] https://www.cs.toronto.edu/~kriz/cifar.html.
[3] https://cs.stanford.edu/~acoates/stl10/.

3.2. Baseline models. VAE-PU-Bayes (Wawrzeńczyk and Mielniczuk, 2024) (abbreviated as VP-B) based on the results of Na *et al*. (2020) was selected as a state-of-the-art PU method working under SAR assumption. Estimation of $s(x)$ is performed using a simple feedforward Neural Network (NN).

3.3. Estimation of $\widetilde{\pi}$. We start by considering estimators of $\widetilde{\pi}$ which will be then incorporated into proposed classifiers. The first estimator is $\widehat{\widetilde{\pi}}$ defined in (7) and is called the direct estimator of $\widetilde{\pi}$ in the following. Alternatively, one can apply classical EM algorithm (see the work of Saerens *et al*. (2002)), to estimate labeled-shifted prior probability $\widetilde{\pi}$ in this setting. For this aim $y(x)$, which in classical case is estimated by, e.g., NN based on fully observable sample $(X_i, Y_i), i = 1, \ldots, n$ is estimated here by one of the PU estimators of posterior probability under selection bias. We use the variational PU-Bayes classifier VP-B (Wawrzeńczyk and Mielniczuk, 2024) for this purpose. For other possible variational classifiers designed for this framework; see the works of Na *et al*. (2020) or Wawrzeńczyk and Mielniczuk (2023). Note that as π it is assumed known, it replaces in the original algorithm fraction of positive observation in the sample, which is not available. The resulting EM estimator is denoted by $\widehat{\widetilde{\pi}}_{EM}$.

3.4. Empirical Bayes rules classifiers. We define several classifiers based derivations in Section 2.1 . In our experiments, we aimed to evaluate the proposed methods and choose the one most adequate to handle the biased label shift PU problems.

CLS estimator. This is an empirical analogue of the Bayes rule defined in (11) with $\widehat{\widetilde{\pi}}$ defined in (7). We use VP-B to estimate $y(x)$ and a separate NN to estimate $s(x)$. The proposal is named the cut-off label shift (CLS) estimator.

CLS-EM estimator. The estimator is defined similarly to CLS, the only difference being that threshold is changed from $\widehat{\widetilde{\pi}}$ to $\widehat{\widetilde{\pi}}_{EM}$.

ALS estimator. We note that due to (4) and (5) we have

$$\widetilde{y}(x) = \frac{\widetilde{s}(x)}{s(x)} \times y(x). \tag{12}$$

Table 2. MSE of direct and EM estimators of $\widetilde{\pi}$ (mean over all datasets).

c	0.1		0.3		0.5		0.7		0.9	
Estimator	Direct	EM	Direct	EM	Direct	EM	Direct	EM	Direct	EM
$\widetilde{\pi}$										
0.1	**0.025**	0.047	**0.014**	0.033	**0.008**	0.027	**0.006**	0.025	**0.003**	0.025
0.3	0.037	**0.029**	0.025	**0.019**	**0.014**	**0.014**	**0.011**	0.015	**0.005**	0.014
0.5	0.042	**0.023**	0.027	**0.017**	0.016	**0.012**	**0.012**	**0.012**	**0.005**	0.009
0.7	0.055	**0.034**	0.030	**0.028**	**0.019**	0.024	**0.014**	0.021	**0.006**	0.018
0.9	0.070	**0.047**	**0.039**	0.041	**0.025**	0.037	**0.018**	0.032	**0.008**	0.029
No shift	0.036	**0.024**	0.021	**0.017**	0.013	**0.012**	**0.010**	**0.010**	**0.004**	0.007
Mean error	0.044	**0.034**	**0.026**	**0.026**	**0.016**	0.021	**0.012**	0.019	**0.005**	0.017

Plugging-in this expression in expression for the classification function $\widetilde{d}_B^{PU}(x)$ we obtain (see (9))

$$\frac{\widetilde{y}(x)-\widetilde{s}(x)}{1-\widetilde{y}(x)}>1 \quad \equiv \quad \frac{\widetilde{s}(x)\left(\frac{y(x)}{s(x)}-1\right)}{1-\frac{\widetilde{s}(x)}{s(x)}y(x)}>1.$$

This gives rise to the competing empirical Bayes rules: estimate $y(x)$ and $s(x)$ based on $\mathcal{D}$ and $\widetilde{s}$ based on $\widetilde{D}$ and apply the formula above with plugged in estimators to construct empirical Bayes rule. We use VP-B to estimate $y(x)$ and separate NNs to estimate $s(x)$ and $\widetilde{s}(x)$. We note that in contrast to the classifiers introduced above one needs to estimate posterior $\widetilde{s}(x)$. We call this estimator the augmented label shift (ALS) estimator.

3.5. General experiment settings. For each experimental setting (i.e., a combination of dataset, label frequency c, target label shift prior $\widetilde{\pi}$ and label shift estimator), we performed 10 experiments, each initialized with a different random seed (equal to experiment number). Data was split between train and test following 70-30 split. Because prediction for labeled examples is trivial in this setting (as $S=1$ implies $Y=1$), instead of using traditional metrics, we define a set of U-metrics. U-metrics are calculated based only on unlabeled stratum of test set, which alleviates trivial prediction impact and puts focus on the classifier performance on the key test subset. As an example, U-accuracy is an accuracy calculated only on the $S=0$ stratum: $\textit{U-ACC} = {n_U}^{-1}\sum_{x_U\in U}\mathbb{I}\{d(x_U,s)=y_U\}$. The whole method and experiment code is available in a public GitHub repository[4].

4. Results of experiments

Table 2 and Fig. 1 contrast the $\widetilde{\pi}$ estimation performance of the direct and EM estimators. Both achieve generally good results, and in both cases, their quality increases as label frequency rises. This is especially apparent for the direct estimator and due to the fact that for low c values the number of labeled samples is low, and thus estimation of $P(S=1)$ in the denominator of (6) becomes sensitive to small deviations. Also, this is consistent with form of the bound in (8), where c appears in the denominator of the bound. Due to this property, EM outperforms the direct estimator up until $c=0.3$, where the direct estimator starts to prevail. As evident in Fig. 1, the direct estimator also tends to perform better for lower shift priors, with EM outperforming direct estimator for the higher prior values. Both this property and the significant variability of results are heavily impacted by the low label frequency performance of both estimators: compare MSEs of both estimators for $c=0.1$ and $c=0.9$ in Table 2. An important thing to note is that using the EM estimator is associated with the additional cost of the iterative EM procedure—it increases the computational complexity of the estimation, making it significantly slower compared to the direct method.

Figure 2 compares the performance of the three estimators proposed in Section 3.4. It is evident that the ALS estimator lags behind the other two methods in terms of U-balanced accuracy. This is likely to be related to its more complex formulation, requiring training of two additional models for the estimation of both $s(x)$ and $\widetilde{s}(x)$. CLS and CLS-EM are significantly more comparable in performance. To help differentiate between them, we report average rank in Table 3, as well as the difference of accuracy between the method and the best classifier in Table 4. Both metrics are averaged over all label frequencies and label shift priors. According to both metrics, CLS outperforms the competitor, with a mean rank of around 1.5 and approximately two times smaller difference from the best estimator. CLS estimator is more consistent over the performed experiments and various datasets, but CLS-EM also has evident benefits in some cases - it often slightly edges out the CLS estimator for high label frequencies. We note also consistently better performance of CLS for $\widetilde{\pi}=0.9$ (the fifth row of Fig. 2). For the case of no shift ($\widetilde{\pi}=\pi$, the last row of Fig. 2) their performance is strikingly similar, with slight superiority

[4] https://github.com/wawrzenczyka/VAE-PU-label-shift.

Fig. 1. Comparison of $\widetilde{\pi}$ estimation errors (MSE) (averaged over all label frequencies).

Table 3. U-balanced accuracy ranks.

Estimator	CIFAR CT	CIFAR VA	MNIST 3v5	MNIST OvE	STL VA	Mean rank
ALS	2.64	2.74	2.72	2.73	2.76	2.72
CLS	**1.61**	1.66	**1.48**	**1.37**	**1.49**	**1.52**
CLS-EM	1.75	**1.60**	1.80	1.89	1.75	1.76

Table 4. U-balanced accuracy difference from best estimator.

Estimator	CIFAR CT	CIFAR VA	MNIST 3v5	MNIST OvE	STL VA	Mean ifference
ALS	0.086	0.089	0.156	0.197	0.142	0.134
CLS	**0.010**	**0.011**	**0.015**	**0.010**	**0.010**	**0.011**
CLS-EM	0.018	0.021	0.033	0.024	0.030	0.025

of CLS-EM for $c \geq 0.7$ in the case of CIFAR VA and STL VA.

Table 5 shows the maximum standard error of the mean U-Balanced accuracy for each estimator, when the maximum is taken over c. The reported errors are small when compared to U-balanced accuracy, which indicates that the results are stable and reliable. The highest standard error is observed for the ALS estimator, which matches the results presented in Fig. 2: the CLS and CLS-EM estimators are more consistent in their performance.

Using MNIST 3v5 as an example, we analyzed how different labeling biases affect the performance of the estimators. While creating the dataset as described in Section 3.1, instead of labeling the most bold examples, we sampled the examples without replacement with step weights. The "standard" labeling reported before corresponds to $0-1$ step weight: the boldest n_L examples are sampled with weight 1, while the remaining have their weights set to 0. Similarly, $0.5-0.5$ step corresponds to the SCAR scenario, as both the top n_L examples and the remaining ones are sampled with equal weight of 0.5. The results are presented in Table 6. We observe that the performance of the estimators is consistent across different labeling biases, with performance generally improving as the datasets get closer to the SCAR scenario (except for the select cases of the ALS estimator), which is obviously the easiest task to solve. This illustrates the robustness of the estimators to different labeling biases regardless of the shift prior.

5. Summary

In this paper, we discuss the issue of label shift phenomenon in the context of augmented PU data. In the general result section, we investigated probabilistic structure of label-shifted augmented PU data, proving that the label shift for general populations carries over to their unlabeled subpopulations. Moreover, we constructed

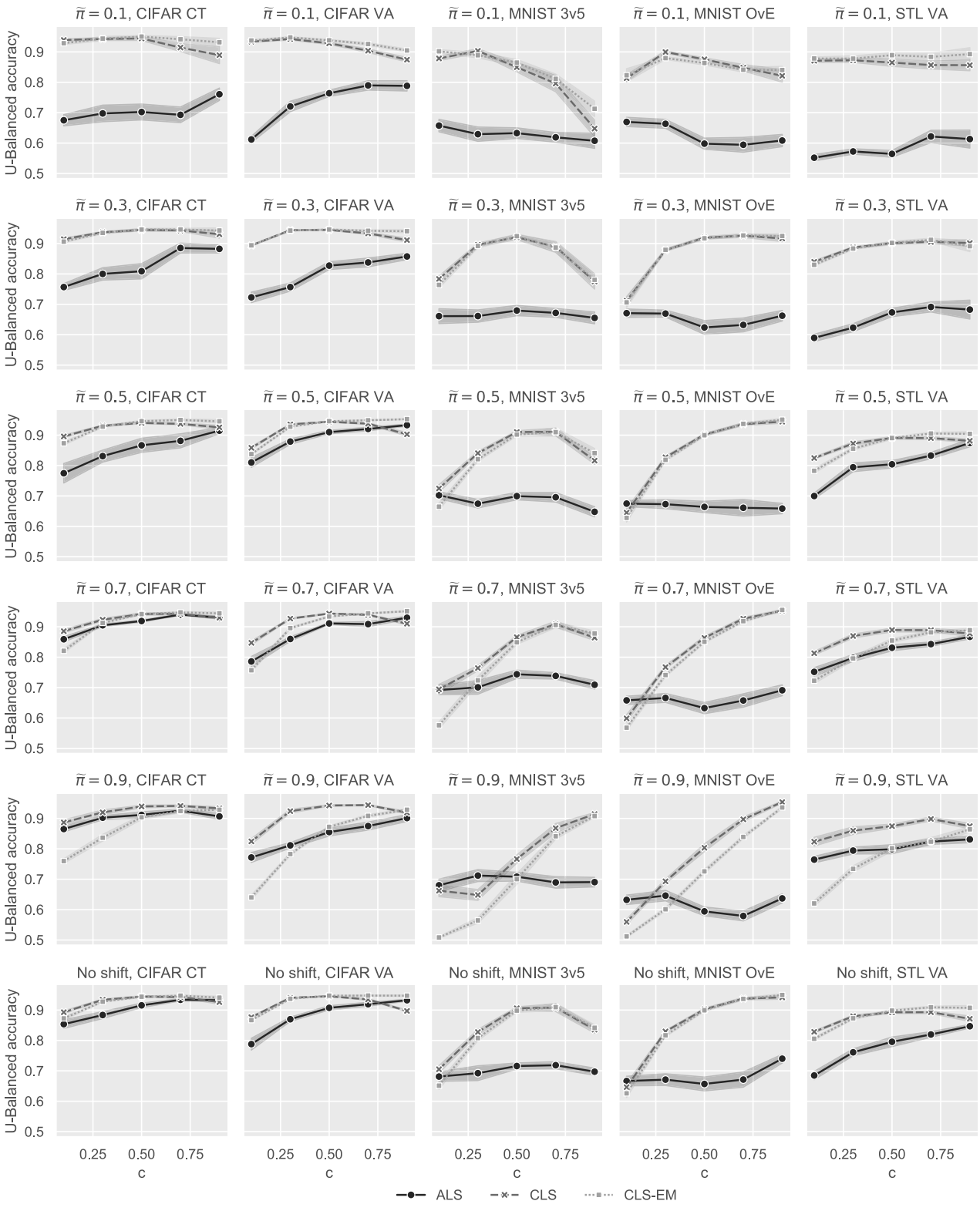

Fig. 2. U-balanced accuracy in label shift scenario for each estimator.

Table 5. Maximum standard error of the mean U-balanced accuracy for each estimator.

Estimator	Dataset	No shift	$\widetilde{\pi}=0.1$	$\widetilde{\pi}=0.3$	$\widetilde{\pi}=0.5$	$\widetilde{\pi}=0.7$	$\widetilde{\pi}=0.9$
ALS	CIFAR CT	0.014	0.027	0.025	0.032	0.008	0.010
	CIFAR VA	0.019	0.017	0.017	0.016	0.016	0.016
	MNIST 3v5	0.024	0.025	0.024	0.016	0.022	0.020
	MNIST OvE	0.025	0.024	0.023	0.027	0.022	0.017
	STL VA	0.016	0.029	0.031	0.014	0.014	0.015
CLS	CIFAR CT	0.007	0.028	0.013	0.006	0.005	0.008
	CIFAR VA	0.003	0.007	0.006	0.005	0.006	0.009
	MNIST 3v5	0.013	0.027	0.025	0.012	0.016	0.019
	MNIST OvE	0.012	0.021	0.013	0.012	0.012	0.013
	STL VA	0.009	0.018	0.011	0.006	0.009	0.014
CLS-EM	CIFAR CT	0.005	0.013	0.008	0.006	0.007	0.011
	CIFAR VA	0.003	0.004	0.005	0.003	0.005	0.007
	MNIST 3v5	0.012	0.025	0.017	0.015	0.013	0.017
	MNIST OvE	0.013	0.021	0.016	0.013	0.009	0.006
	STL VA	0.007	0.021	0.018	0.007	0.009	0.013

Table 6. Mean U-balanced accuracy for MNIST 3v5 with various step probabilities and shift priors, averaged by label frequency.

Estimator	Step size	No shift	$\widetilde{\pi}=0.1$	$\widetilde{\pi}=0.3$	$\widetilde{\pi}=0.5$	$\widetilde{\pi}=0.7$	$\widetilde{\pi}=0.9$
ALS	0 – 1 (standard)	0.701	0.629	0.666	0.684	0.717	0.696
	0.1 – 0.9	0.717	0.576	0.585	0.568	0.739	0.716
	0.3 – 0.7	0.747	0.583	0.576	0.555	0.781	0.745
	0.5 – 0.5 (SCAR)	0.760	0.577	0.587	0.551	0.775	0.745
CLS	0 – 1 (standard)	0.837	0.815	0.853	0.841	0.819	0.772
	0.1 – 0.9	0.889	0.867	0.898	0.885	0.878	0.855
	0.3 – 0.7	0.913	0.908	0.920	0.909	0.902	0.890
	0.5 – 0.5 (SCAR)	0.919	0.918	0.930	0.916	0.914	0.909
CLS-EM	0 – 1 (standard)	0.822	0.836	0.850	0.828	0.787	0.704
	0.1 – 0.9	0.865	0.894	0.906	0.872	0.817	0.720
	0.3 – 0.7	0.890	0.933	0.926	0.896	0.850	0.759
	0.5 – 0.5 (SCAR)	0.900	0.940	0.931	0.905	0.865	0.784

the correct Bayesian rule for the label-shifted sample showing that it is Bayes classifier for the augmented PU data with appropriately modified threshold. Our findings led us to propose three potential classifiers built upon the state-of-the-art VAE-PU-Bayes method: ALS, CLS and CLS-EM. As an intermediate step, we also consider the problem of $\widetilde{\pi}$ estimation, which is a key component of the proposed classifiers; we propose the direct and EM estimators in order to solve this problem. In the experiment section, we show that both the direct and EM estimators perform well in terms of $\widetilde{\pi}$ estimation. We also conclude that the CLS estimator generally outperforms the competing methods, and CLS-EM is a viable alternative in high label frequency scenarios. Future research in this area might investigate the possibility of creating even more stable label shift prior estimators, as well as label shift detection. Moreover, allowing for different propensity scores for training and target sample is of both theoretical and practical interest.

References

Bekker, J. and Davis, J. (2020). Learning from positive and unlabeled data: A survey, *Machine Learning* **109**: 719–760, DOI: 10.1007/s10994-020-05877-5.

Elkan, C. (2001). The foundations of cost-sensitive learning, *Proceedings of the 17th International Joint Conference on Artificial Intelligence, Seattle, USA*, pp. 973–978.

Garg, S., Wu, Y., Balakrishnan, S. and Lipton, Z.C. (2020). A unified view of label shift estimation, *Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, Canada*, pp. 1–11.

González, P., Castaño, A., Chawla, N. and Coz, J. (2017). A review on quantification learning, *ACM Computing Surveys* **50**(5): 1-74.

Iyer, A., Nath, S. and Sarawagi, S. (2014). Maximum mean discrepancy for class ratio estimation: Convergence bounds and kernel selection, *Proceedings of the 31st International Conferencce on Machine Learning, Beijing, China*, pp. 230–238.

Kiryo, R., Niu, G., du Plessis, M.C. and Sugiyama, M. (2017). Positive-unlabeled learning with non-negative risk estimator, *Proceedings of the International Conference on Neural Information Processing Systems, NIPS'17, Long Beach, USA*, pp. 1674–1684.

Lipton, Z.C., Wang, Y. and Smola, A.J. (2018). Detecting and correcting for label shift with black box predictors, *Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden*, pp. 3128–3136.

Mielniczuk, J. (1985). Estimation of number of errors in case of repetitive quality control, *Probability and Mathematical Statistics* **6**(2): 131–136.

Na, B., Kim, H., Song, K., Joo, W., Kim, Y.-Y. and Moon, I. (2020). Deep generative positive-unlabeled learning under selection bias, *Proceedings of CIKM'20, New York, NY, USA*, pp. 1155–1164.

Saerens, M., Latinne, P. and Decaestecker, C. (2002). Adjusting the outputs of a classifier to new a priori probabilities: A simple procedure, *Neural Computing* **14**(1): 21–41.

Vaz, A., Izbicki, R. and Stern, R. (2019). Quantification under prior probability shift: The ratio estimator and its extensions, *Journal of Machine Learning Research* **20**: 1–33.

Wainwright, M. (2019). *High-dimensional Statistics*, Cambridge University Press, Cambridge.

Wawrzeńczyk, A. and Mielniczuk, J. (2022). Revisiting strategies for fitting logistic regression for positive and unlabeled data, *International Journal of Applied Mathematics and Computer Science* **32**(2): 299–309, DOI: 10.34768/amcs-2022-0022.

Wawrzeńczyk, A. and Mielniczuk, J. (2023). One-class classification approach to variational learning from biased positive unlabelled data, *Proceedings of the European Conference on Artificial Intelligence, ECAI'23, Cracow, Poland*, pp. 1720–1727.

Wawrzeńczyk, A. and Mielniczuk, J. (2024). Augmented prediction of a true class for positive unlabeled data under selection bias, *Proceedings of the European Conference on Artificial Intelligence, ECAI'24, Santiago de Compostela, Spain*, pp. 2725–2733.

Ye, C., Tsuchida, R., Petersson, L. and Barnes, N. (2024). Label shift estimation for class-imbalance problem: A Bayesian approach, *IEEE/CVF 2024, Seattle, USA*.

Jan Mielniczuk is a full professor at the Institute of Computer Science, Polish Academy of Sciences, and a professor at the Faculty of Mathematics and Information Sciences, Warsaw University of Technology. His main research contributions concern computational statistics and data mining, in particular time series modeling and prediction, inference for high dimensional and misspecified data, model selection, computer-intensive methods, asymptotic analysis, and quantification of dependence. He is an author and co-author of two books and over 90 articles.

Adam Wawrzeńczyk holds an MSc degree in computer science from the Faculty of Mathematics and Information Sciences, Warsaw University of Technology, and is currently a PhD student at the Doctoral School of Information and Biomedical Technologies of the Polish Academy of Sciences. His research interests include recent advances in machine and deep learning, in particular inference from partially observable data. He specializes in creating software utilizing machine learning methods for practical applications, including forecasting and LLMs.

Received: 2 December 2024
Revised: 19 March 2025
Accepted: 11 May 2025