Supporting Information: Verifying the Selected Completely at Random Assumption in Positive-Unlabeled Learning

Paweł Teisseyre[a,b,*], **Konrad Furmańczyk**[c] **and Jan Mielniczuk**[a,b]

[a]Polish Academy of Sciences, Warsaw, Poland
[b]Warsaw University of Technology, Warsaw, Poland
[c]Warsaw University of Life Sciences, Warsaw, Poland

1 Proofs

1.1 Proof of equation (1) in the main paper

First, using the Law of Total Probability and the fact that in PU learning negative observations cannot be labeled, i.e., $P(S=1|X=x, Y=0)=0$ and $P(S=1|Y=0)=0$ we have

$$P(S=1|X=x)=e(x)P(Y=1|X=x)$$
$$P(S=1)=c\cdot P(Y=1)$$

Applying Bayes' theorem twice and the above equations we have

$$\begin{aligned}
P(X=x|S=1) &= \frac{P(S=1|X=x)P(X=x)}{P(S=1)} \\
&= \frac{e(x)P(Y=1|X=x)P(X=x)}{c\cdot P(Y=1)} \\
&= \frac{e(x)}{c}\frac{P(X=x|Y=1)P(Y=1)P(X=x)}{P(Y=1)P(X=x)} \\
&= \frac{e(x)}{c}P(X=x|Y=1)
\end{aligned}$$

The above equality proves that $e(x)=c$ (SCAR) is equivalent to $P_{X|S=1}=P_{X|Y=1}$.

We also note that under SCAR, variables X and S are conditionally independent given Y and thus $P_{X|S=0,Y=1}=P_{X|Y=1}$ and $P_{X|S=0,Y=0}=P_{X|Y=0}$ [1]. Thus $P_{X|S=0}$ is mixture of positive and negative distribution but, in contrast to P_X, with proportions $\pi(1-c)/(1-\pi c)$ and $(1-\pi)/(1-\pi c)$, respectively.

1.2 Proof of Theorem 2 in the main paper

Part (ii) is obvious in view of (4) in the main paper. To prove part (iii), it is enough to show that

$$\mathrm{Cov}(Y_{[k]},Y_{[l]})=P(Y_{[k]}=1,Y_{[l]}=1)-P(Y_{[k]}=1)P(Y_{[l]}=1)$$

is non-negative. However, we have for $f_{kl}(x,z)$ denoting joint density of kth and lth order statistic in the sequence $(\widetilde{y}(X_i))_i$

$$\begin{aligned}
\mathrm{Cov}(Y_{[k]},Y_{[l]}) &= \int (h(z)-EY_{[k]})(h(x)-EY_{[l]})f_{kl}(x,z)\,dxdz \\
&= \mathrm{Cov}(h(\widetilde{y}(X)_{(k)}),h(\widetilde{y}(X)_{(l)})),
\end{aligned}$$

where (4) from the main paper is used for the last equality. Moreover, it follows from the proof of part (i) that $h(\widetilde{y}(X)_{(k)})$ is kth order statistic in the sequence $h(\widetilde{y}(X_1)),\dots h(\widetilde{y}(X_m))$. The result then follows from [2], Lemma 2.1 (a).

1.3 Proof of Theorem 3 in the main paper

Observe that

$$\begin{aligned}
P(Y_{[1]}=1,Y_{[2]}=1,\dots,Y_{[k]}=1) &= E(\prod_{i=1}^{k}Y_{[i]}) \\
&= E(\prod_{i=1}^{k}h(\widetilde{y}(X)_{(i)}))=E(\prod_{i=1}^{k}F^{-1}(U_{(i)})),
\end{aligned}$$

where $U_{(i)}$ is $(m-i+1)$th order statistics from the uniform distribution on $(0,1)$. From the equation (5) in the main paper ($F^{-1}(z)\geq z$), we have

$$E(\prod_{i=1}^{k}F^{-1}(U_{(i)}))\geq E(\prod_{i=1}^{k}U_{(i)}).$$

Since $E(U_{(i)})=\frac{m-i+1}{m+1}=1-\frac{i}{m+1}$ [3], then to prove Theorem 2 it is enough to show

$$E(\prod_{i=1}^{k}U_{(i)})\geq\prod_{i=1}^{k}E(U_{(i)}). \tag{1}$$

In this proof, we use an auxiliary lemma

Lemma (i) $E(\prod_{i=2}^{k}U_{(i)}|U_{(1)}=t)$ is an increasing function of t. (ii) $Cov(U_{(1)},\prod_{i=2}^{k}U_{(i)})\geq 0$.

From Lemma (ii) we obtain

$$Cov(U_{(1)},\prod_{i=2}^{k}U_{(i)})=E((U_{(1)}-E(U_{(1)}))\prod_{i=2}^{k}U_{(i)})\geq 0.$$

* Corresponding Author. Email: teisseyrep@ipipan.waw.pl

Hence, we have

$$E(\prod_{i=1}^{k} U_{(i)}) \geq E(U_{(1)})E(\prod_{i=2}^{k} U_{(i)}).$$

By the induction wrt k, we have (1) after noting that the result is valid for $k = 2$ in view of Lemma 2.1 (a) in [2].

The proof of Lemma (i) is straightforward and follows from standard integration or the fact that given $U_{(1)} = t$ $U_{(2)}, \ldots U_{(k)}$ are order statistics from uniform distribution $U[0, t]$ on $[0, t]$ and $U[0, t]$ $U[0, s]$ are stochastically ordered. Now, we will show Lemma (ii). Note that

$$\begin{aligned} Cov(U_{(1)}, \prod_{i=2}^{k} U_{(i)}) &= E(U_{(1)}(\prod_{i=2} U_{(i)} - E(\prod_{i=2}^{k} U_{(i)}))) \\ &= E(U_{(1)}(E(\prod_{i=2}^{k} U_{(i)}|U_{(1)}) - E(\prod_{i=2}^{k} U_{(i)}))). \end{aligned}$$

Let $G(t) := E(\prod_{i=2}^{k} U_{(i)}|U_{(1)} = t) - E(\prod_{i=2}^{k} U_{(i)})$. From Lemma (i) $G(t)$ is increasing function of t, then reasoning as in the proof of Lemma 2.1 in Bickel [2] we have that there exists a constant C such that $G(U_{(1)})(U_{(1)} - C) \geq 0$. Obviously, this implies that

$$\begin{aligned} Cov(U_{(1)}, \prod_{i=2}^{k} U_{(i)}) &= E(U_{(1)}G(U_{(1)})) \\ &= E(G(U_{(1)})(U_{(1)} - C)) \geq 0. \end{aligned}$$

Therefore, we obtain Lemma (ii).

1.4 Remarks on function $h(z)$

In the main paper we consider function

$$h(z) = P(Y = 1|\widetilde{y}(X) = z, S = 0).$$

Let us denote by $\pi' = P(Y = 1|S = 0)$ and

$$A = P(\widetilde{y}(X) = z|Y = 1, S = 0)\pi'$$

$$B = (1 - \pi')P(\widetilde{y}(X) = z|Y = 0, S = 0).$$

Then we have

$$\begin{aligned} h(z) &= \frac{p(\widetilde{y}(X) = z|Y = 1, S = 0)\pi'}{A + B} \\ &= \frac{p(\widetilde{y}(X) = z|Y = 1)\pi'}{p(\widetilde{y}(X) = z|Y = 1)\pi' + (1 - \pi')p(\widetilde{y}(X) = z|Y = 0)}, \end{aligned}$$

where the equality above follows from the fact that under SCAR, variables X and S are conditionally independent given Y and thus $P_{\widetilde{y}(X)|S=0,Y=i} = P_{\widetilde{y}(X)|Y=i}$ for $i = 0, 1$. In particular, the fact that $h(z)$ is strictly increasing, is equivalent to the ratio of densities of $\widetilde{y}(X)$ for positive and negative observations being increasing.

2 Experiments

2.1 Datasets

Tabular datasets with multiple classes were transformed into binary classification datasets, such that the positive class includes the most common class, and the remaining classes are combined into the negative class. In the case of image datasets, we define the binary class variable depending on the particular dataset. In the USPS positive digits less than five constitute the positive class, and remaining digits constitute the negative class. In Fashion, clothing items worn on the upper body are marked as positive cases, the remaining images are in the negative class. In CIFAR10, all machines form a positive class and animals form a negative class. For image data, pre-trained deep neural network Resnet18 was used to extract the feature vector of size 512, which is defined as an outcome of the average pooling layer. For datasets having more than 50 features, we applied simple PU filter considered in [4] and we selected the 50 most informative features. The method is based on calculating the mutual information between each feature and label indicator S.

Table 1. Statistics of the considered data sets.

Dataset	n	p	$P(Y=1)$	positives	type
Breast-w	699	9	0.34	241.0	tabular
Wdbc	569	30	0.37	212.0	tabular
Banknote	1372	4	0.44	610.0	tabular
Segment	2310	19	0.14	330.0	tabular
CIFAR10	50000	-	0.4	20000	images
USPS	7291	-	0.58	4240	images
Fashion	60000	-	0.5	30000	images

Table 2. Type I error (probability of rejecting H_0 when H_0 is true, also called observed level of significance) for $c = 0.3$. Cases in which the type I error exceeds the assumed level $\alpha = 0.05$ are marked in red.

Dataset	KL	KLCOV	KS	NB AUC
Breast	0.03 ± 0.02	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01
Wdbc	0.08 ± 0.03	0.1 ± 0.03	0.01 ± 0.01	0.01 ± 0.01
Banknote	0.04 ± 0.02	1.0 ± 0.0	0.02 ± 0.01	0.05 ± 0.03
Segment	0.11 ± 0.03	0.3 ± 0.05	0.01 ± 0.01	0.04 ± 0.02
CIFAR10	0.04 ± 0.02	0.01 ± 0.01	0.0 ± 0.0	0.0 ± 0.0
USPS	0.02 ± 0.01	0.08 ± 0.03	0.01 ± 0.01	0.02 ± 0.01
Fashion	0.05 ± 0.02	0.01 ± 0.01	0.0 ± 0.0	0.0 ± 0.0

Table 3. Type I error (probability of rejecting H_0 when H_0 is true, also called observed level of significance) for $c = 0.7$. Cases in which the type I error exceeds the assumed level $\alpha = 0.05$ are marked in red.

Dataset	KL	KLCOV	KS	NB AUC
Breast	0.01 ± 0.01	0.0 ± 0.0	0.01 ± 0.01	0.01 ± 0.01
Wdbc	0.08 ± 0.03	0.23 ± 0.04	0.01 ± 0.01	0.0 ± 0.0
Banknote	0.01 ± 0.01	0.88 ± 0.03	0.03 ± 0.02	0.04 ± 0.02
Segment	0.08 ± 0.03	0.13 ± 0.03	0.05 ± 0.02	0.02 ± 0.01
CIFAR10	0.14 ± 0.03	0.0 ± 0.0	0.05 ± 0.01	0.01 ± 0.01
USPS	0.03 ± 0.02	0.03 ± 0.02	0.02 ± 0.01	0.02 ± 0.01
Fashion	0.0 ± 0.0	0.0 ± 0.0	0.01 ± 0.01	0.01 ± 0.01

Table 4. Power of the tests (probability of rejecting H_0 when H_1 is true) for $c = 0.5$ and labelling schemes S1, S2 and S3.

Labelling scheme S1					
Dataset	g	KL	KLCOV	KS	NB AUC
Breast	1	0.15 ± 0.04	0.02 ± 0.01	0.24 ± 0.04	0.02 ± 0.01
	2	0.79 ± 0.04	0.17 ± 0.04	0.74 ± 0.04	0.54 ± 0.05
Wdbc	1	0.79 ± 0.04	0.3 ± 0.05	0.91 ± 0.03	0.8 ± 0.04
	2	1.0 ± 0.0	0.66 ± 0.05	1.0 ± 0.0	1.0 ± 0.0
Banknote	1	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Segment	1	0.15 ± 0.04	0.19 ± 0.04	0.4 ± 0.05	0.27 ± 0.04
	2	0.18 ± 0.04	0.26 ± 0.04	0.95 ± 0.02	0.83 ± 0.04
CIFAR10	1	0.07 ± 0.03	0.0 ± 0.0	0.41 ± 0.05	0.0 ± 0.0
	2	0.19 ± 0.04	0.01 ± 0.01	0.57 ± 0.05	0.0 ± 0.0
USPS	1	0.96 ± 0.02	0.92 ± 0.03	0.87 ± 0.03	0.92 ± 0.03
	2	0.96 ± 0.02	0.9 ± 0.03	0.94 ± 0.02	0.91 ± 0.03
Fashion	1	0.99 ± 0.01	0.95 ± 0.02	0.99 ± 0.01	1.0 ± 0.0
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0

Labelling scheme S2					
Dataset	g	KL	KLCOV	KS	NB AUC
Breast	1	0.95 ± 0.02	0.37 ± 0.05	0.95 ± 0.02	0.89 ± 0.03
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Wdbc	1	0.59 ± 0.05	0.52 ± 0.05	0.2 ± 0.04	0.15 ± 0.04
	2	0.89 ± 0.03	0.66 ± 0.05	0.79 ± 0.04	0.65 ± 0.05
Banknote	1	0.8 ± 0.04	1.0 ± 0.0	0.46 ± 0.05	1.0 ± 0.0
	2	1.0 ± 0.0	1.0 ± 0.0	0.99 ± 0.01	1.0 ± 0.0
Segment	1	0.06 ± 0.02	0.13 ± 0.03	0.06 ± 0.02	0.21 ± 0.04
	2	0.12 ± 0.03	0.16 ± 0.04	0.3 ± 0.05	0.79 ± 0.04
CIFAR10	1	0.29 ± 0.05	0.2 ± 0.04	0.05 ± 0.02	0.08 ± 0.03
	2	0.9 ± 0.03	0.88 ± 0.03	0.83 ± 0.04	0.79 ± 0.04
USPS	1	0.48 ± 0.05	0.48 ± 0.05	0.44 ± 0.05	0.37 ± 0.05
	2	1.0 ± 0.0	1.0 ± 0.0	0.99 ± 0.01	0.98 ± 0.01
Fashion	1	0.53 ± 0.05	0.41 ± 0.05	0.65 ± 0.05	0.26 ± 0.04
	2	0.99 ± 0.01	0.98 ± 0.01	1.0 ± 0.0	1.0 ± 0.0

Labelling scheme S3					
Dataset	g	KL	KLCOV	KS	NB AUC
Breast	1	1.0 ± 0.0	0.94 ± 0.02	1.0 ± 0.0	1.0 ± 0.0
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Wdbc	1	0.76 ± 0.04	0.56 ± 0.05	0.45 ± 0.05	0.34 ± 0.05
	2	0.97 ± 0.02	0.91 ± 0.03	0.99 ± 0.01	0.97 ± 0.02
Banknote	1	1.0 ± 0.0	1.0 ± 0.0	0.88 ± 0.03	1.0 ± 0.0
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Segment	1	0.12 ± 0.03	0.15 ± 0.04	0.17 ± 0.04	0.49 ± 0.05
	2	0.11 ± 0.03	0.18 ± 0.04	0.5 ± 0.05	0.96 ± 0.02
CIFAR10	1	0.63 ± 0.05	0.55 ± 0.05	0.3 ± 0.05	0.39 ± 0.05
	2	1.0 ± 0.0	0.98 ± 0.01	1.0 ± 0.0	0.97 ± 0.02
USPS	1	0.88 ± 0.03	0.81 ± 0.04	0.76 ± 0.04	0.74 ± 0.04
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Fashion	1	0.8 ± 0.04	0.79 ± 0.04	0.93 ± 0.03	0.7 ± 0.05
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0

Table 5. Power of the tests (probability of rejecting H_0 when H_1 is true) for $c = 0.3$ and labelling schemes S1, S2 and S3.

Labelling scheme S1					
Dataset	g	KL	KLCOV	KS	NB AUC
Breast	1	0.2 ± 0.04	0.0 ± 0.0	0.27 ± 0.04	0.04 ± 0.02
	2	0.82 ± 0.04	0.16 ± 0.04	0.81 ± 0.04	0.4 ± 0.05
Wdbc	1	0.93 ± 0.03	0.05 ± 0.02	0.91 ± 0.03	0.92 ± 0.03
	2	1.0 ± 0.0	0.21 ± 0.04	1.0 ± 0.0	1.0 ± 0.0
Banknote	1	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Segment	1	0.3 ± 0.05	0.43 ± 0.05	0.4 ± 0.05	0.21 ± 0.04
	2	0.38 ± 0.05	0.52 ± 0.05	0.88 ± 0.03	0.72 ± 0.04
CIFAR10	1	0.01 ± 0.01	0.01 ± 0.01	0.0 ± 0.0	0.02 ± 0.01
	2	0.05 ± 0.02	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01
USPS	1	0.96 ± 0.02	0.94 ± 0.02	0.91 ± 0.03	0.92 ± 0.03
	2	0.91 ± 0.03	0.96 ± 0.02	0.91 ± 0.03	0.94 ± 0.02
Fashion	1	1.0 ± 0.0	0.97 ± 0.02	1.0 ± 0.0	0.99 ± 0.01
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0

Labelling scheme S2					
Dataset	g	KL	KLCOV	KS	NB AUC
Breast	1	0.96 ± 0.02	0.29 ± 0.05	0.9 ± 0.03	0.72 ± 0.04
	2	1.0 ± 0.0	0.98 ± 0.01	1.0 ± 0.0	1.0 ± 0.0
Wdbc	1	0.3 ± 0.05	0.08 ± 0.03	0.11 ± 0.03	0.15 ± 0.04
	2	0.6 ± 0.05	0.17 ± 0.04	0.47 ± 0.05	0.58 ± 0.05
Banknote	1	0.82 ± 0.04	1.0 ± 0.0	0.32 ± 0.05	0.97 ± 0.02
	2	0.98 ± 0.01	1.0 ± 0.0	0.71 ± 0.05	1.0 ± 0.0
Segment	1	0.22 ± 0.04	0.24 ± 0.04	0.04 ± 0.02	0.25 ± 0.04
	2	0.31 ± 0.05	0.42 ± 0.05	0.3 ± 0.05	0.8 ± 0.04
CIFAR10	1	0.54 ± 0.05	0.33 ± 0.05	0.18 ± 0.04	0.08 ± 0.03
	2	0.97 ± 0.02	0.97 ± 0.02	0.98 ± 0.01	0.75 ± 0.04
USPS	1	0.69 ± 0.05	0.65 ± 0.05	0.63 ± 0.05	0.47 ± 0.05
	2	1.0 ± 0.0	0.99 ± 0.01	0.99 ± 0.01	1.0 ± 0.0
Fashion	1	0.65 ± 0.05	0.37 ± 0.05	0.63 ± 0.05	0.19 ± 0.04
	2	0.98 ± 0.01	1.0 ± 0.0	1.0 ± 0.0	0.98 ± 0.01

Labelling scheme S3					
Dataset	g	KL	KLCOV	KS	NB AUC
Breast	1	0.85 ± 0.04	0.22 ± 0.04	0.83 ± 0.04	0.63 ± 0.05
	2	1.0 ± 0.0	0.91 ± 0.03	1.0 ± 0.0	1.0 ± 0.0
Wdbc	1	0.26 ± 0.04	0.14 ± 0.03	0.1 ± 0.03	0.13 ± 0.03
	2	0.38 ± 0.05	0.25 ± 0.04	0.41 ± 0.05	0.43 ± 0.05
Banknote	1	0.32 ± 0.05	1.0 ± 0.0	0.13 ± 0.03	0.85 ± 0.04
	2	0.86 ± 0.03	1.0 ± 0.0	0.46 ± 0.05	1.0 ± 0.0
Segment	1	0.26 ± 0.04	0.41 ± 0.05	0.03 ± 0.02	0.17 ± 0.04
	2	0.27 ± 0.04	0.45 ± 0.05	0.36 ± 0.05	0.74 ± 0.04
CIFAR10	1	0.5 ± 0.05	0.24 ± 0.04	0.19 ± 0.04	0.06 ± 0.02
	2	0.96 ± 0.02	0.83 ± 0.04	0.94 ± 0.02	0.68 ± 0.05
USPS	1	0.72 ± 0.04	0.52 ± 0.05	0.54 ± 0.05	0.38 ± 0.05
	2	1.0 ± 0.0	0.97 ± 0.02	0.99 ± 0.01	0.97 ± 0.02
Fashion	1	0.65 ± 0.05	0.29 ± 0.05	0.49 ± 0.05	0.18 ± 0.04
	2	0.96 ± 0.02	0.91 ± 0.03	0.99 ± 0.01	0.95 ± 0.02

Table 6. Power of the tests (probability of rejecting H_0 when H_1 is true) for $c = 0.7$ and labelling schemes S1, S2 and S3.

Labelling scheme S1					
Dataset	g	KL	KLCOV	KS	NB AUC
Breast	1	0.11 ± 0.03	0.0 ± 0.0	0.17 ± 0.04	0.01 ± 0.01
	2	0.61 ± 0.05	0.02 ± 0.01	0.66 ± 0.05	0.38 ± 0.05
Wdbc	1	0.67 ± 0.05	0.67 ± 0.05	0.72 ± 0.04	0.49 ± 0.05
	2	0.98 ± 0.01	0.95 ± 0.02	1.0 ± 0.0	1.0 ± 0.0
Banknote	1	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Segment	1	0.23 ± 0.04	0.26 ± 0.04	0.32 ± 0.05	0.11 ± 0.03
	2	0.31 ± 0.05	0.43 ± 0.05	0.89 ± 0.03	0.63 ± 0.05
CIFAR10	1	0.51 ± 0.05	0.0 ± 0.0	0.73 ± 0.04	0.05 ± 0.02
	2	0.56 ± 0.05	0.01 ± 0.01	0.89 ± 0.03	0.08 ± 0.03
USPS	1	0.94 ± 0.02	0.96 ± 0.02	0.96 ± 0.02	0.93 ± 0.03
	2	0.93 ± 0.03	0.9 ± 0.03	0.92 ± 0.03	0.94 ± 0.02
Fashion	1	0.89 ± 0.03	0.71 ± 0.05	0.98 ± 0.01	0.86 ± 0.03
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Labelling scheme S2					
Dataset	g	KL	KLCOV	KS	NB AUC
Breast	1	0.88 ± 0.03	0.22 ± 0.04	0.93 ± 0.03	0.75 ± 0.04
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Wdbc	1	0.65 ± 0.05	0.59 ± 0.05	0.2 ± 0.04	0.1 ± 0.03
	2	0.96 ± 0.02	0.92 ± 0.03	0.82 ± 0.04	0.6 ± 0.05
Banknote	1	0.83 ± 0.04	1.0 ± 0.0	0.56 ± 0.05	1.0 ± 0.0
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Segment	1	0.13 ± 0.03	0.47 ± 0.05	0.06 ± 0.02	0.18 ± 0.04
	2	0.15 ± 0.04	0.79 ± 0.04	0.19 ± 0.04	0.73 ± 0.04
CIFAR10	1	0.08 ± 0.03	0.03 ± 0.02	0.01 ± 0.01	0.21 ± 0.04
	2	0.42 ± 0.05	0.44 ± 0.05	0.35 ± 0.05	0.75 ± 0.04
USPS	1	0.23 ± 0.04	0.24 ± 0.04	0.31 ± 0.05	0.25 ± 0.04
	2	0.91 ± 0.03	0.9 ± 0.03	0.97 ± 0.02	0.93 ± 0.03
Fashion	1	0.19 ± 0.04	0.13 ± 0.03	0.34 ± 0.05	0.11 ± 0.03
	2	0.9 ± 0.03	0.86 ± 0.03	1.0 ± 0.0	0.93 ± 0.03
Labelling scheme S3					
Dataset	g	KL	KLCOV	KS	NB AUC
Breast	1.0	0.89 ± 0.03	0.22 ± 0.04	0.94 ± 0.02	0.79 ± 0.04
	2.0	1.0 ± 0.0	0.92 ± 0.03	1.0 ± 0.0	1.0 ± 0.0
Wdbc	1.0	0.56 ± 0.05	0.7 ± 0.05	0.22 ± 0.04	0.1 ± 0.03
	2.0	0.98 ± 0.01	0.91 ± 0.03	0.81 ± 0.04	0.54 ± 0.05
Banknote	1.0	0.84 ± 0.04	1.0 ± 0.0	0.47 ± 0.05	1.0 ± 0.0
	2.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Segment	1.0	0.09 ± 0.03	0.44 ± 0.05	0.06 ± 0.02	0.13 ± 0.03
	2.0	0.11 ± 0.03	0.77 ± 0.04	0.11 ± 0.03	0.71 ± 0.05
CIFAR10	1.0	0.07 ± 0.03	0.05 ± 0.02	0.01 ± 0.01	0.27 ± 0.04
	2.0	0.47 ± 0.05	0.41 ± 0.05	0.07 ± 0.03	0.44 ± 0.05
USPS	1.0	0.16 ± 0.04	0.41 ± 0.05	0.35 ± 0.05	0.29 ± 0.05
	2.0	0.9 ± 0.03	0.88 ± 0.03	0.88 ± 0.03	0.89 ± 0.03
Fashion	1.0	0.24 ± 0.04	0.12 ± 0.03	0.44 ± 0.05	0.12 ± 0.03
	2.0	0.91 ± 0.03	0.78 ± 0.04	0.99 ± 0.01	0.77 ± 0.04

Table 7. The probability of rejecting H_0 for S1 and KS test statistic. Cases in which the probability of rejecting H_0 when it is true exceeds the assumed level $\alpha = 0.05$ are marked in red, while cases in which $\alpha = 0.05$ is not exceeded are marked in blue.

Dataset	g	True π	Overestimated π ($\hat{\pi} = \pi + 0.2\pi$)	Underestimated π ($\hat{\pi} = \pi - 0.2\pi$)
Breast	0	0.0 ± 0.0	0.28 ± 0.04	0.0 ± 0.0
	1	0.24 ± 0.04	0.86 ± 0.03	0.02 ± 0.01
	2	0.74 ± 0.04	1.0 ± 0.0	0.23 ± 0.04
Wdbc	0	0.0 ± 0.0	0.04 ± 0.02	0.0 ± 0.0
	1	0.91 ± 0.03	1.0 ± 0.0	0.3 ± 0.05
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Banknote	0	0.04 ± 0.02	0.16 ± 0.04	0.0 ± 0.0
	1	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
Segment	0	0.02 ± 0.01	0.04 ± 0.02	0.0 ± 0.0
	1	0.49 ± 0.05	0.66 ± 0.05	0.13 ± 0.03
	2	0.93 ± 0.03	0.94 ± 0.02	0.77 ± 0.04
CIFAR10	0	0.05 ± 0.02	0.0 ± 0.0	0.44 ± 0.05
	1	0.41 ± 0.05	0.01 ± 0.01	0.83 ± 0.04
	2	0.57 ± 0.05	0.11 ± 0.03	0.84 ± 0.04
USPS	0	0.02 ± 0.01	0.15 ± 0.04	0.05 ± 0.03
	1	0.87 ± 0.03	0.95 ± 0.02	0.94 ± 0.02
	2	0.94 ± 0.02	0.97 ± 0.02	0.96 ± 0.02
Fashion	0	0.02 ± 0.01	0.25 ± 0.04	0.01 ± 0.01
	1	0.99 ± 0.01	1.0 ± 0.0	0.55 ± 0.05
	2	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0

References

[1] J. Bekker and J. Davis. Learning from positive and unlabeled data: a survey. *Machine Learning*, 109:719–760, 2020.

[2] P. Bickel. Some contributions to theory of order statistics. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1*, pages 575–591, 1967.

[3] P. Bickel and K. Doksum. *Mathematical Statistics*. CRC, London, 2015.

[4] K. Sechidis and G. Brown. Simple strategies for semi-supervised feature selection. *Machine Learning*, 107(2):357–395, 2018.