Supplementary Material for "Semi-Supervised Classification Based on Classification from Positive and Unlabeled Data"

Tomoya Sakai [1 2] **Marthinus Christoffel du Plessis** **Gang Niu** [1] **Masashi Sugiyama** [2 1]

A. Proofs of Theorems

In this section, we give the proofs of Theorems in Section 4.

A.1. Proof of Theorem 1

Recall that

$$\begin{aligned} R^{\gamma}_{\text{N-PUNU}}(g) &= (1-\gamma)R_{\text{N-PU}}(g) + \gamma R_{\text{N-NU}}(g) \\ &= (2-2\gamma)\theta_{\text{P}}R_{\text{P}}(g) + 2\gamma\theta_{\text{N}}R_{\text{N}}(g) + (1-\gamma)R_{\text{U,N}}(g) + \gamma R_{\text{U,P}}(g) + \text{Const}, \\ R^{\gamma}_{\text{N-PNPU}}(g) &= (1-\gamma)R_{\text{PN}}(g) + \gamma R_{\text{N-PU}}(g) \\ &= (1+\gamma)\theta_{\text{P}}R_{\text{P}}(g) + (1-\gamma)\theta_{\text{N}}R_{\text{N}}(g) + \gamma R_{\text{U,N}}(g) + \text{Const}, \\ R^{\gamma}_{\text{N-PNNU}}(g) &= (1-\gamma)R_{\text{PN}}(g) + \gamma R_{\text{N-NU}}(g) \\ &= (1-\gamma)\theta_{\text{P}}R_{\text{P}}(g) + (1+\gamma)\theta_{\text{N}}R_{\text{N}}(g) + \gamma R_{\text{U,P}}(g) + \text{Const}. \end{aligned}$$

Let $\widehat{R}_{\text{P}}(g)$, $\widehat{R}_{\text{N}}(g)$, $\widehat{R}_{\text{U,P}}(g)$ and $\widehat{R}_{\text{U,N}}(g)$ be the empirical risks. In order to prove Theorem 1, the following concentration lemma is needed:

Lemma 1 *For any $\delta > 0$, we have these uniform deviation bounds with probability at least $1-\delta/3$:*

$$\begin{aligned} \sup_{g\in\mathcal{G}}(R_{\text{P}}(g) - \widehat{R}_{\text{P}}(g)) &\le \frac{C_w C_\phi}{\sqrt{n_{\text{P}}}} + \sqrt{\frac{\ln(3/\delta)}{2n_{\text{P}}}}, \\ \sup_{g\in\mathcal{G}}(R_{\text{N}}(g) - \widehat{R}_{\text{N}}(g)) &\le \frac{C_w C_\phi}{\sqrt{n_{\text{N}}}} + \sqrt{\frac{\ln(3/\delta)}{2n_{\text{N}}}}, \\ \sup_{g\in\mathcal{G}}(R_{\text{U,P}}(g) - \widehat{R}_{\text{U,P}}(g)) &\le \frac{C_w C_\phi}{\sqrt{n_{\text{U}}}} + \sqrt{\frac{\ln(3/\delta)}{2n_{\text{U}}}}, \\ \sup_{g\in\mathcal{G}}(R_{\text{U,N}}(g) - \widehat{R}_{\text{U,N}}(g)) &\le \frac{C_w C_\phi}{\sqrt{n_{\text{U}}}} + \sqrt{\frac{\ln(3/\delta)}{2n_{\text{U}}}}. \end{aligned}$$

All inequalities in Lemma 1 are from the basic *uniform deviation bound* using the Rademacher complexity (Mohri et al., 2012), *Talagrand's contraction lemma* (Ledoux & Talagrand, 1991), as well as the fact that the Lipschitz constant of ℓ_{R} is $1/2$. For these reasons, the detailed proof of Lemma 1 is omitted.

[1]The University of Tokyo, Japan [2]RIKEN, Japan. Correspondence to: Tomoya Sakai <sakai@ms.k.u-tokyo.ac.jp>.

Proceedings of the 34^{th} International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017.

Consider $R^{\gamma}_{\text{N-PNPU}}(g)$. It is clear that

$$\begin{aligned}&\sup_{g\in\mathcal{G}}(R^{\gamma}_{\text{N-PNPU}}(g)-\widehat{R}^{\gamma}_{\text{N-PNPU}}(g))\\&\le(1+\gamma)\theta_{\mathrm{P}}\sup_{g\in\mathcal{G}}(R_{\mathrm{P}}(g)-\widehat{R}_{\mathrm{P}}(g))+(1-\gamma)\theta_{\mathrm{N}}\sup_{g\in\mathcal{G}}(R_{\mathrm{N}}(g)-\widehat{R}_{\mathrm{N}}(g))+\gamma\sup_{g\in\mathcal{G}}(R_{\mathrm{U,N}}(g)-\widehat{R}_{\mathrm{U,N}}(g)).\end{aligned}$$

Therefore, by applying Lemma 1, for any $\delta>0$, it holds with probability at least $1-\delta$ that

$$\sup_{g\in\mathcal{G}}(R^{\gamma}_{\text{N-PNPU}}(g)-\widehat{R}^{\gamma}_{\text{N-PNPU}}(g))\le\frac{1}{2}C_{w,\phi,\delta}\cdot\chi(1+\gamma,1-\gamma,\gamma).$$

Since $I(g)\le 2R^{\gamma}_{\text{N-PNPU}}$, with the same probability,

$$\sup_{g\in\mathcal{G}}(I(g)-2\widehat{R}^{\gamma}_{\text{N-PNPU}}(g))\le C_{w,\phi,\delta}\cdot\chi(1+\gamma,1-\gamma,\gamma).$$

Similarly, $\sup_{g\in\mathcal{G}}(I(g)-2\widehat{R}^{\gamma}_{\text{N-PNNU}}(g))\le C_{w,\phi,\delta}\cdot\chi(1-\gamma,1+\gamma,\gamma)$ with probability at least $1-\delta$.

Finally, $R^{\gamma}_{\text{N-PUNU}}(g)$ is slightly more involved, for that there are both $R_{\mathrm{U,P}}(g)$ and $R_{\mathrm{U,N}}(g)$. From $\ell_{\mathrm{R}}(m)+\ell_{\mathrm{R}}(-m)=1$, we can know $R_{\mathrm{U,P}}(g)+R_{\mathrm{U,N}}(g)=1$ and then

$$(1-\gamma)R_{\mathrm{U,N}}(g)+\gamma R_{\mathrm{U,P}}(g)=\begin{cases}(2\gamma-1)R_{\mathrm{U,P}}(g)+\mathrm{Const} & \gamma\ge 1/2,\\(1-2\gamma)R_{\mathrm{U,N}}(g)+\mathrm{Const} & \gamma<1/2.\end{cases}$$

As a result, $\sup_{g\in\mathcal{G}}(I(g)-2\widehat{R}^{\gamma}_{\text{N-PUNU}}(g))\le C_{w,\phi,\delta}\cdot\chi(2-2\gamma,2\gamma,|2\gamma-1|)$ with probability at least $1-\delta$.

A.2. Proof of Theorem 2

In fact,

$$\ell_{\mathrm{TS}}(m)=\begin{cases}1/4 & m\le 0,\\(m-1)^2/4 & 0<m\le 1,\\0 & m>1,\end{cases}$$

and after plugging this $\ell_{\mathrm{TS}}(m)$ into $\widetilde{\ell}_{\mathrm{TS}}(m)$,

$$\begin{aligned}\widetilde{\ell}_{\mathrm{TS}}(m)&=\ell_{\mathrm{TS}}(m)-\ell_{\mathrm{TS}}(-m)\\&=\begin{cases}1/4 & m\le -1,\\1/4-(m+1)^2/4 & -1<m\le 0,\\(m-1)^2/4-1/4 & 0<m\le 1,\\-1/4 & m>1.\end{cases}\end{aligned}$$

It is easy to see that $\ell_{\mathrm{TS}}(m)$ and $\widetilde{\ell}_{\mathrm{TS}}(m)$ are Lipschitz continuous with the same Lipschitz constant $1/2$.

Next, recall that

$$\begin{aligned}R^{\gamma}_{\text{C-PUNU}}(g)&=(1-\gamma)R_{\text{C-PU}}(g)+\gamma R_{\text{C-NU}}(g)\\&=(1-\gamma)\theta_{\mathrm{P}}R'_{\mathrm{P}}(g)+\gamma\theta_{\mathrm{N}}R'_{\mathrm{N}}(g)+(1-\gamma)R_{\mathrm{U,N}}(g)+\gamma R_{\mathrm{U,P}}(g),\\R^{\gamma}_{\text{C-PNPU}}(g)&=(1-\gamma)R_{\mathrm{PN}}(g)+\gamma R_{\text{C-PU}}(g)\\&=(1-\gamma)\theta_{\mathrm{P}}R_{\mathrm{P}}(g)+(1-\gamma)\theta_{\mathrm{N}}R_{\mathrm{N}}(g)+\gamma\theta_{\mathrm{P}}R'_{\mathrm{P}}(g)+\gamma R_{\mathrm{U,N}}(g),\\R^{\gamma}_{\text{C-PNNU}}(g)&=(1-\gamma)R_{\mathrm{PN}}(g)+\gamma R_{\text{C-NU}}(g)\\&=(1-\gamma)\theta_{\mathrm{P}}R_{\mathrm{P}}(g)+(1-\gamma)\theta_{\mathrm{N}}R_{\mathrm{N}}(g)+\gamma\theta_{\mathrm{N}}R'_{\mathrm{N}}(g)+\gamma R_{\mathrm{U,P}}(g).\end{aligned}$$

Let $\widehat{R}_{\mathrm{P}}(g)$, $\widehat{R}_{\mathrm{N}}(g)$, $\widehat{R}_{\mathrm{U,P}}(g)$, $\widehat{R}_{\mathrm{U,N}}(g)$, $\widehat{R}'_{\mathrm{P}}(g)$ and $\widehat{R}'_{\mathrm{N}}(g)$ be the empirical risks. Again, the following concentration lemma is needed:

Lemma 2 *For any $\delta > 0$, we have these uniform deviation bounds with probability at least $1 - \delta/4$:*

$$\sup_{g\in\mathcal{G}}(R_{\mathrm{P}}(g) - \widehat{R}_{\mathrm{P}}(g)) \leq \frac{C_w C_\phi}{\sqrt{n_{\mathrm{P}}}} + \sqrt{\frac{\ln(4/\delta)}{32 n_{\mathrm{P}}}},$$

$$\sup_{g\in\mathcal{G}}(R_{\mathrm{N}}(g) - \widehat{R}_{\mathrm{N}}(g)) \leq \frac{C_w C_\phi}{\sqrt{n_{\mathrm{N}}}} + \sqrt{\frac{\ln(4/\delta)}{32 n_{\mathrm{N}}}},$$

$$\sup_{g\in\mathcal{G}}(R_{\mathrm{U,P}}(g) - \widehat{R}_{\mathrm{U,P}}(g)) \leq \frac{C_w C_\phi}{\sqrt{n_{\mathrm{U}}}} + \sqrt{\frac{\ln(4/\delta)}{32 n_{\mathrm{U}}}},$$

$$\sup_{g\in\mathcal{G}}(R_{\mathrm{U,N}}(g) - \widehat{R}_{\mathrm{U,N}}(g)) \leq \frac{C_w C_\phi}{\sqrt{n_{\mathrm{U}}}} + \sqrt{\frac{\ln(4/\delta)}{32 n_{\mathrm{U}}}},$$

$$\sup_{g\in\mathcal{G}}(R'_{\mathrm{P}}(g) - \widehat{R}'_{\mathrm{P}}(g)) \leq \frac{C_w C_\phi}{\sqrt{n_{\mathrm{P}}}} + \sqrt{\frac{\ln(4/\delta)}{8 n_{\mathrm{P}}}},$$

$$\sup_{g\in\mathcal{G}}(R'_{\mathrm{N}}(g) - \widehat{R}'_{\mathrm{N}}(g)) \leq \frac{C_w C_\phi}{\sqrt{n_{\mathrm{N}}}} + \sqrt{\frac{\ln(4/\delta)}{8 n_{\mathrm{N}}}}.$$

The detailed proof of Lemma 2 is omitted for the same reason as Lemma 1. The difference is due to that $0 \leq \ell_{\mathrm{TS}}(m) \leq 1/4$ and $-1/4 \leq \widetilde{\ell}_{\mathrm{TS}}(m) \leq 1/4$ whereas $0 \leq \ell_{\mathrm{R}}(m) \leq 1$ just like $0 \leq \ell_{0\text{-}1}(m) \leq 1$. For convenience, we will relax $1/32$ to $1/8$ in the square root for $R_{\mathrm{P}}(g)$, $R_{\mathrm{N}}(g)$, $R_{\mathrm{U,P}}(g)$, $R_{\mathrm{U,N}}(g)$.

Consider $R^{\gamma}_{\text{C-PUNU}}(g)$. By applying Lemma 2, for any $\delta > 0$, it holds with probability at least $1 - \delta$ that

$$\sup_{g\in\mathcal{G}}(R^{\gamma}_{\text{C-PUNU}}(g) - \widehat{R}^{\gamma}_{\text{C-PUNU}}(g)) \leq \frac{1}{4} C'_{w,\phi,\delta} \cdot \chi(1-\gamma, \gamma, 1).$$

Since $I(g) \leq 4R^{\gamma}_{\text{C-PUNU}}$, with the same probability,

$$\sup_{g\in\mathcal{G}}(I(g) - 4\widehat{R}^{\gamma}_{\text{C-PUNU}}(g)) \leq C'_{w,\phi,\delta} \cdot \chi(1-\gamma, \gamma, 1).$$

The other two generalization error bounds can be proven similarly.

A.3. Proofs of Theorems 3 and 4

Note that g is independent of the data for evaluating $\widehat{R}^{\gamma}_{\text{N-PUNU}}(g)$, since it is fixed in the evaluation. Thus, $\mathrm{Var}_{\mathrm{P}}[\widehat{R}_{\mathrm{P}}(g)] = \sigma^2_{\mathrm{P}}(g)/n_{\mathrm{P}}$ and $\mathrm{Var}_{\mathrm{N}}[\widehat{R}_{\mathrm{N}}(g)] = \sigma^2_{\mathrm{N}}(g)/n_{\mathrm{N}}$. When $n_{\mathrm{U}} \to \infty$,

$$\begin{aligned}
\mathrm{Var}[\widehat{R}^{\gamma}_{\text{N-PUNU}}(g)] &= 4(1-\gamma)^2\theta^2_{\mathrm{P}} \,\mathrm{Var}_{\mathrm{P}}[\widehat{R}_{\mathrm{P}}(g)] + 4\gamma^2\theta^2_{\mathrm{N}} \,\mathrm{Var}_{\mathrm{N}}[\widehat{R}_{\mathrm{N}}(g)] \\
&= 4(1-\gamma)^2\psi_{\mathrm{P}} + 4\gamma^2\psi_{\mathrm{N}} \\
&= 4(\psi_{\mathrm{P}} + \psi_{\mathrm{N}})\gamma^2 - 8\psi_{\mathrm{P}}\gamma + 4\psi_{\mathrm{P}},
\end{aligned}$$

and it is obvious that $\gamma_{\text{N-PUNU}} \in [0, 1]$. All other claims in Theorem 3 follow from that $\mathrm{Var}[\widehat{R}^{\gamma}_{\text{N-PUNU}}(g)]$ is quadratic in γ, that $\mathrm{Var}[\widehat{R}^{\gamma}_{\text{N-PUNU}}(g)] = \mathrm{Var}[\widehat{R}_{\mathrm{PN}}(g)]$ at $\gamma = 1/2$, and that $\gamma_{\text{N-PUNU}} < 1/2$ if $\psi_{\mathrm{P}} < \psi_{\mathrm{N}}$ or $\gamma_{\text{N-PUNU}} > 1/2$ if $\psi_{\mathrm{P}} > \psi_{\mathrm{N}}$.

Likewise, when $n_{\mathrm{U}} \to \infty$,

$$\begin{aligned}
\mathrm{Var}[\widehat{R}^{\gamma}_{\text{N-PNPU}}(g)] &= (1+\gamma)^2\psi_{\mathrm{P}} + (1-\gamma)^2\psi_{\mathrm{N}}, \\
\mathrm{Var}[\widehat{R}^{\gamma}_{\text{N-PNNU}}(g)] &= (1-\gamma)^2\psi_{\mathrm{P}} + (1+\gamma)^2\psi_{\mathrm{N}},
\end{aligned}$$

and $\gamma_{\text{N-PNPU}} \geq 0$ if $\psi_{\mathrm{P}} \leq \psi_{\mathrm{N}}$ or $\gamma_{\text{N-PNNU}} \geq 0$ if $\psi_{\mathrm{P}} \geq \psi_{\mathrm{N}}$. The rest of proof of Theorem 4 is analogous to that of Theorem 3.

B. Experimental Setting

Here, we summarized the experimental settings.

B.1. Implementation in Our Experiments

We implemented the ER by ourselves, and for the other methods, we used the codes available at the authors' websites:

- LapSVM: `http://www.dii.unisi.it/~melacci/lapsvmp/`
- SMIR: `http://www.ms.k.u-tokyo.ac.jp/software/SMIR.zip`
- WellSVM: `http://lamda.nju.edu.cn/code_WellSVM.ashx`
- S4VM: `http://lamda.nju.edu.cn/files/s4vm.rar`.

Note that we modified the original code of the S4VM for transductive learning to inductive learning according to Li & Zhou (2015).

B.2. Parameter Candidates in Our Experiments

The regularization parameters for all the methods were chosen from $\{10^{-5}, 10^{-4}, \ldots, 10^{2}\}$, except the regularization parameter of the SMIR for the squared loss mutual information (SMI) and that of the S4VM for labeled data. The number of nearest-neighbors to construct Laplacian matrix for the LapSVM was chosen from the candidates $\{5, 6, \ldots, 10\}$. The combination parameter η of PNU classification was chosen from $\{-1, -0.9, \ldots, 1\}$, and γ of PUNU classification was chosen from $\{0, 0.05, \ldots, 1\}$. We chose these hyper-parameters by five-fold cross-validation. The parameter for the ℓ_2-regularizer of the SMIR is set at $\gamma_{\mathrm{S}}/(n \cdot \min_{k \in \{\pm 1\}} p(y=k)) + 0.001$, where γ_S is the regularization parameter for the SMI. The regularization parameter of the S4VM for the labeled data is set at 1. The other parameters were set at the default values.

B.3. Data Set Description of Image Classification Data Set

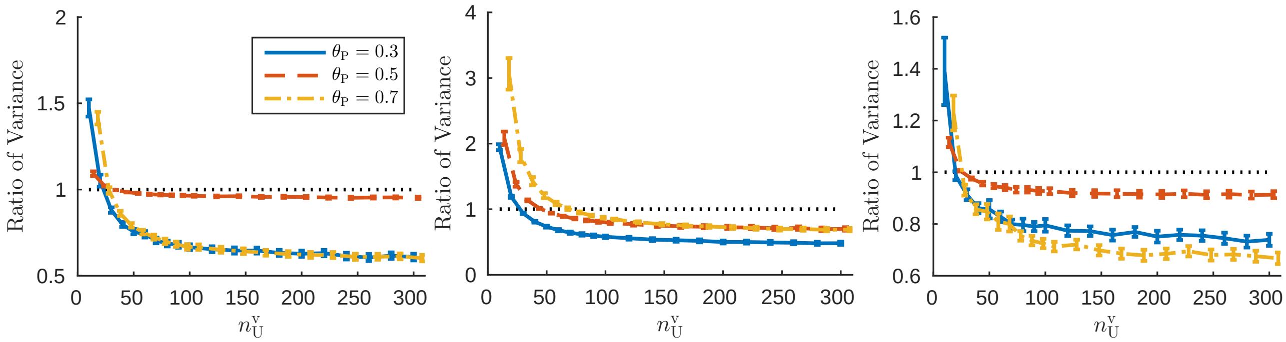
Table 1 is the description of the data sets used in the image classification experiment.

Table 1. The description of the data set used in the image classification experiment.

Data set	Data sources	#Samples
Arts	Art Gallery vs. Art Studio	$(m_{\mathrm{P}} = 15000)$ $(m_{\mathrm{N}} = 15000)$
Deserts	Desert Sand vs. Desert Vegetation	$(m_{\mathrm{P}} = 15000)$ $(m_{\mathrm{N}} = 5556)$
Fields	Field Wild vs. Field Cultivated	$(m_{\mathrm{P}} = 15000)$ $(m_{\mathrm{N}} = 8117)$
Stadiums	Stadium Baseball vs. Stadium Football	$(m_{\mathrm{P}} = 15000)$ $(m_{\mathrm{N}} = 15000)$
Platforms	Subway Station vs. Train Station	$(m_{\mathrm{P}} = 5597)$ $(m_{\mathrm{N}} = 15000)$
Temples	Temple East Asia vs. Temple South Asia	$(m_{\mathrm{P}} = 8691)$ $(m_{\mathrm{N}} = 7178)$

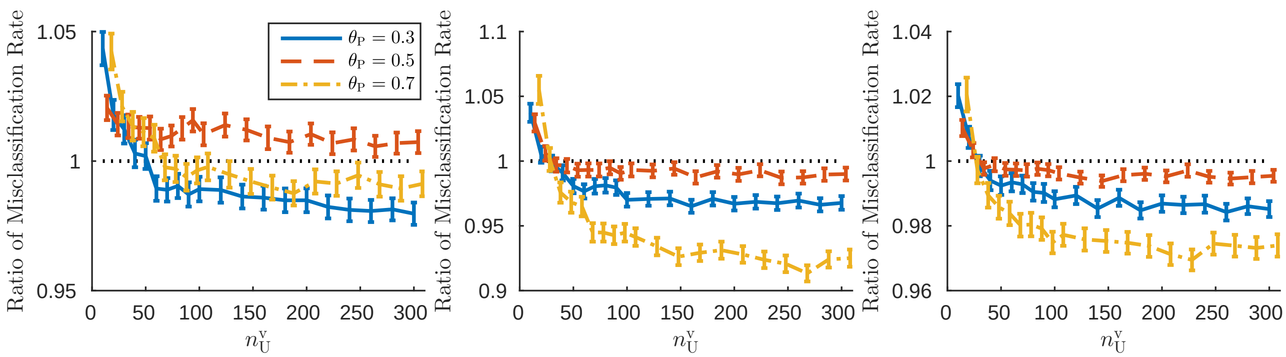
C. Supplementary Results for Experimental Analyses

Figure 1 and Figure 2 respectively show the results of variance reduction and comparison of validation scores. The details of experimental setting and the interpretation of results can be found in Section 5.1.



(a) Banana ($d=2$) (b) Waveform ($d=21$) (c) Splice ($d=60$)

Figure 1. Average and standard error of the ratio between the variance of the empirical PNU risk and that of the PN risk, $\mathrm{Var}[\widehat{R}^{\eta}_{\mathrm{PNU}}(\widehat{g}_{\mathrm{PN}})]/\,\mathrm{Var}[\widehat{R}_{\mathrm{PN}}(\widehat{g}_{\mathrm{PN}})]$, as a function of the number of unlabeled samples over 100 trials. Although the variance reduction is proved for an infinite number of samples, it can be observed with a finite number of samples.



(a) Banana ($d=2$) (b) Waveform ($d=21$) (c) Splice ($d=60$)

Figure 2. Average and standard error of the ratio between the misclassification rate of $\widehat{g}^{\mathrm{PNU}}_{\mathrm{PN}}$ and that of $\widehat{g}^{\mathrm{PN}}_{\mathrm{PN}}$ as a function of unlabeled samples over 1000 trials. In many cases, the ratio becomes less than 1 or at worst almost 1, implying that the PNU risk is a promising alternative to the standard PN risk in validation if unlabeled data is available.

D. Magnified Versions of Experimental Results

Here, we show magnified versions of the experimental results in Section 5.

Table 2. Magnified version of Table 1: Average and standard error of the misclassification rates of each method over 50 trials for benchmark data sets. Boldface numbers denote the best and comparable methods in terms of average misclassifications rate according to a t-test at a significance level of 5%. The bottom row gives the number of best/comparable cases of each method.

Data set	n_{L}	PNU	PUNU	ER	LapSVM	SMIR	WellSVM	S4VM
Banana	10	**30.1 (1.0)**	**32.1 (1.1)**	35.8 (1.0)	36.9 (1.0)	37.7 (1.1)	41.8 (0.6)	45.3 (1.0)
$d = 2$	50	**19.0 (0.6)**	26.4 (1.2)	20.6 (0.7)	21.3 (0.7)	21.1 (1.0)	42.6 (0.5)	38.7 (0.9)
Phoneme	10	32.5 (0.8)	33.5 (1.0)	33.4 (1.2)	36.5 (1.5)	36.4 (1.2)	**28.4 (0.6)**	33.7 (1.4)
$d = 5$	50	28.1 (0.5)	32.8 (0.9)	27.8 (0.6)	27.0 (0.8)	28.6 (1.0)	26.8 (0.4)	**25.1 (0.2)**
Magic	10	**31.7 (0.8)**	34.1 (0.9)	34.2 (1.1)	37.9 (1.3)	36.0 (1.2)	**30.1 (0.8)**	33.3 (0.9)
$d = 10$	50	**29.9 (0.8)**	33.4 (0.9)	30.9 (0.5)	31.0 (0.9)	**30.8 (0.9)**	**28.8 (0.8)**	**29.2 (0.4)**
Image	10	**29.8 (0.9)**	**31.7 (0.8)**	33.7 (1.1)	36.6 (1.2)	36.7 (1.2)	34.7 (1.1)	35.9 (1.0)
$d = 18$	50	**20.7 (0.8)**	26.6 (1.1)	**20.8 (0.8)**	**20.3 (1.0)**	**20.9 (0.9)**	27.2 (1.0)	23.2 (0.7)
Susy	10	**44.6 (0.6)**	**45.0 (0.6)**	47.7 (0.4)	48.2 (0.4)	**45.1 (0.7)**	48.0 (0.3)	46.8 (0.3)
$d = 18$	50	**38.9 (0.6)**	41.5 (0.6)	**37.9 (0.7)**	43.1 (0.6)	43.9 (0.8)	43.8 (0.7)	42.1 (0.4)
German	10	**40.8 (0.9)**	**42.4 (0.7)**	43.6 (0.9)	45.9 (0.7)	46.2 (0.8)	**42.4 (0.8)**	**42.0 (0.7)**
$d = 20$	50	**36.2 (0.8)**	39.0 (0.8)	38.9 (0.6)	40.6 (0.6)	38.4 (1.1)	38.5 (1.0)	**34.9 (0.5)**
Waveform	10	**17.4 (0.6)**	**18.0 (0.9)**	18.5 (0.6)	24.9 (1.4)	**18.0 (1.0)**	**16.7 (0.6)**	20.8 (0.8)
$d = 21$	50	16.3 (0.6)	23.7 (1.2)	**14.2 (0.4)**	18.1 (0.8)	**15.4 (0.6)**	15.5 (0.5)	15.3 (0.3)
ijcnn1	10	43.6 (0.6)	**40.3 (1.0)**	49.7 (0.1)	49.2 (0.3)	44.0 (1.0)	45.9 (0.7)	49.3 (0.8)
$d = 22$	50	**34.5 (0.8)**	37.1 (0.9)	**35.5 (0.8)**	**33.4 (1.1)**	49.4 (0.3)	46.2 (0.8)	48.6 (0.4)
g50c	10	11.4 (0.6)	12.5 (0.6)	23.3 (2.3)	39.8 (1.6)	21.9 (1.3)	**6.6 (0.4)**	27.0 (1.4)
$d = 50$	50	12.5 (1.1)	10.1 (0.6)	8.7 (0.4)	22.5 (1.5)	10.6 (0.6)	**7.4 (0.4)**	12.1 (0.5)
covtype	10	**46.2 (0.4)**	**46.0 (0.4)**	**46.0 (0.5)**	47.1 (0.5)	47.9 (0.5)	**46.9 (0.6)**	**46.4 (0.4)**
$d = 54$	50	**41.3 (0.5)**	42.3 (0.5)	**41.0 (0.4)**	**41.5 (0.5)**	46.2 (0.8)	43.6 (0.6)	**40.8 (0.4)**
Spambase	10	27.2 (0.9)	28.1 (1.1)	31.8 (1.4)	39.7 (1.4)	30.9 (1.3)	**23.8 (0.8)**	36.1 (1.5)
$d = 57$	50	23.4 (1.0)	26.6 (1.0)	22.1 (0.7)	28.5 (1.3)	20.9 (0.5)	**19.1 (0.4)**	24.5 (0.9)
Splice	10	**38.3 (0.8)**	**39.3 (0.8)**	43.9 (0.8)	47.9 (0.5)	41.6 (0.7)	42.0 (1.0)	42.4 (0.6)
$d = 60$	50	**30.6 (0.8)**	34.7 (0.9)	**30.9 (0.8)**	38.8 (1.0)	**30.6 (0.9)**	40.9 (0.8)	35.9 (0.7)
phishing	10	**24.2 (1.2)**	**25.8 (1.0)**	**27.3 (1.6)**	37.2 (1.6)	27.6 (1.6)	27.5 (1.4)	31.7 (1.3)
$d = 68$	50	**15.8 (0.6)**	18.3 (0.8)	**15.4 (0.5)**	21.1 (1.3)	**14.7 (0.8)**	17.2 (0.7)	16.7 (0.8)
a9a	10	**31.4 (0.9)**	**31.3 (1.0)**	34.3 (1.2)	41.0 (1.1)	37.3 (1.3)	**33.1 (1.2)**	34.3 (1.2)
$d = 83$	50	27.9 (0.6)	29.9 (0.8)	28.6 (0.7)	33.3 (1.0)	**26.9 (0.7)**	28.9 (0.8)	**26.2 (0.4)**
Coil2	10	**38.7 (0.8)**	**40.1 (0.8)**	42.8 (0.7)	43.9 (0.8)	43.2 (0.8)	**39.1 (0.9)**	44.0 (0.8)
$d = 241$	50	**23.2 (0.6)**	30.5 (0.9)	**23.6 (0.9)**	**22.8 (0.9)**	25.1 (0.9)	**22.6 (0.8)**	25.4 (0.8)
w8a	10	**35.9 (0.9)**	**33.6 (1.0)**	41.6 (1.0)	46.6 (0.8)	39.4 (0.9)	42.1 (0.8)	43.0 (0.8)
$d = 300$	50	**28.1 (0.7)**	**27.6 (0.6)**	**27.0 (0.9)**	38.7 (0.8)	**28.0 (0.9)**	33.7 (0.8)	35.2 (1.0)
#Best/Comp.		23	13	11	4	9	13	7

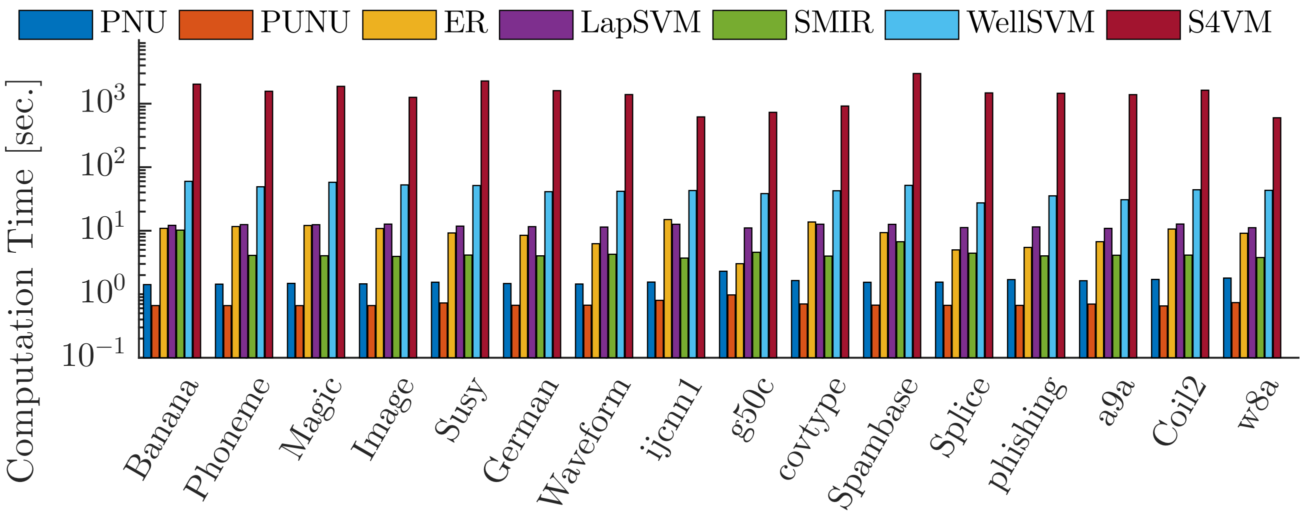


Figure 3. Magnified version of Figure 3: Average computation time over 50 trials for benchmark data sets when $n_{\mathrm{L}} = 50$.

Table 3. Magnified version of Table 2: Average and standard error of misclassification rates over 30 trials for the Places 205 data set. Boldface numbers denote the best and comparable methods in terms of the average misclassification rate according to a t-test at a significance level of 5%.

Data set	n_{U}	θ_{P}	$\widehat{\theta}_{\mathrm{P}}$	PNU	ER	LapSVM	SMIR	WellSVM
Arts	1000	0.50	0.49 (0.01)	**27.4 (1.3)**	**26.6 (0.5)**	**26.1 (0.7)**	40.1 (3.9)	**27.5 (0.5)**
	5000	0.50	0.50 (0.01)	**24.8 (0.6)**	26.1 (0.5)	26.1 (0.4)	30.1 (1.6)	N/A
	10000	0.50	0.52 (0.01)	**25.6 (0.7)**	**25.4 (0.5)**	**25.5 (0.6)**	N/A	N/A
Deserts	1000	0.73	0.67 (0.01)	**13.0 (0.5)**	15.3 (0.6)	16.7 (0.8)	17.2 (0.8)	18.2 (0.7)
	5000	0.73	0.67 (0.01)	**13.4 (0.4)**	**13.3 (0.5)**	16.6 (0.6)	24.4 (0.6)	N/A
	10000	0.73	0.68 (0.01)	**13.3 (0.5)**	**13.7 (0.6)**	16.8 (0.8)	N/A	N/A
Fields	1000	0.65	0.57 (0.01)	**22.4 (1.0)**	26.2 (1.0)	26.6 (1.3)	28.2 (1.1)	26.6 (0.8)
	5000	0.65	0.57 (0.01)	**20.6 (0.5)**	22.6 (0.6)	24.7 (0.8)	29.6 (1.2)	N/A
	10000	0.65	0.57 (0.01)	**21.6 (0.6)**	**22.5 (0.6)**	25.0 (0.9)	N/A	N/A
Stadiums	1000	0.50	0.50 (0.01)	**11.4 (0.4)**	**11.5 (0.5)**	12.5 (0.5)	**17.4 (3.6)**	**11.7 (0.4)**
	5000	0.50	0.50 (0.01)	**11.0 (0.5)**	**10.9 (0.3)**	**11.1 (0.3)**	13.4 (0.7)	N/A
	10000	0.50	0.51 (0.00)	**10.7 (0.3)**	**10.9 (0.3)**	**11.2 (0.2)**	N/A	N/A
Platforms	1000	0.27	0.33 (0.01)	**21.8 (0.5)**	23.9 (0.6)	24.1 (0.5)	30.1 (2.3)	26.2 (0.8)
	5000	0.27	0.34 (0.01)	**23.3 (0.8)**	**24.4 (0.7)**	**24.9 (0.7)**	26.6 (0.3)	N/A
	10000	0.27	0.34 (0.01)	**21.4 (0.5)**	24.3 (0.6)	24.8 (0.5)	N/A	N/A
Temples	1000	0.55	0.51 (0.01)	**43.9 (0.7)**	**43.9 (0.6)**	**43.4 (0.6)**	50.7 (1.6)	**44.3 (0.5)**
	5000	0.55	0.54 (0.01)	**43.4 (0.9)**	**43.0 (0.6)**	**43.1 (1.0)**	**43.6 (0.7)**	N/A
	10000	0.55	0.50 (0.01)	**45.2 (0.8)**	**44.4 (0.8)**	**44.2 (0.7)**	N/A	N/A

References

Ledoux, M. and Talagrand, M. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, 1991.

Li, Y.-F. and Zhou, Z.-H. Towards making unlabeled data never hurt. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1):175–188, 2015.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. *Foundations of Machine Learning*. MIT Press, 2012.