Constant False Alarm Rate for Online One-class SVM Learning

Yongjian Xue, Pierre Beauseroy
Institut Charles Delaunay (ICD)/LM2S - CNRS
Université de Champagne, Université de Technologie de Troyes, Troyes, France
yongjian.xue@utt.fr, pierre.beauseroy@utt.fr

utt université de technologie Troyes
Institut Charles Delaunay
LM2S

STMR
UMR 6281

Motivation

- : Existing online SVM is based on C-SVM, which is not suitable for one-class SVM.
- : For C-OCSVM, the false alarm rate decreases and the miss alarm rate increases as more samples are added.
- : The objective of this paper is to develop an online one-class SVM learning algorithm with stable performance.

The main problem

Let $\mathcal{A}_n = \{\mathbf{x}_i, i = 1, ..., n\}$ be training data, when new sample $\mathbf{x}_{n+1}$ is added,

- **The ν one-class SVM can be rewritten as**:

$$\min_{\boldsymbol{w},\boldsymbol{\xi},\rho} \frac{(n+\gamma)\nu}{2}\|\boldsymbol{w}\|^2 - (n+\gamma)\nu\rho + \sum_{i=1}^{n}\xi_i + \gamma\xi_{n+1} \quad (1)$$
$$s.t.\ \ \langle \boldsymbol{w}, \phi(\boldsymbol{x}_i)\rangle \geq \rho - \xi_i, \xi_i \geq 0, \rho \geq 0, i = 1, ..., n+1,$$

where γ increases from 0 to 1, which indicates the online learning from n to $n+1$ samples, and the Lagrangian parameters $\boldsymbol{\alpha}^{\gamma}$ will be piecewise linear with γ.

- **Dual problem**:

$$\min_{\boldsymbol{\alpha}^{\gamma}} \frac{1}{2(n+\gamma)\nu}\sum_{i,j}\alpha_i^{\gamma}\alpha_j^{\gamma}K(\boldsymbol{x}_i, \boldsymbol{x}_j) \quad (2)$$
$$s.t. \begin{cases} 0 \leq \alpha_i^{\gamma} \leq 1, i = 1, ..., n, 0 \leq \alpha_{n+1}^{\gamma} \leq \gamma, \\ \sum_{i=1}^{n+1}\alpha_i^{\gamma} = (n+\gamma)\nu. \end{cases}$$

- **Lagrangian index groups**:
  - $\mathcal{M} = \{i : f^{\gamma}(\mathbf{x}_i) = 0, \alpha_i^{\gamma} \in [0, 1], i = 1, ..., n\}\bigcup\{n+1 : f^{\gamma}(\mathbf{x}_{n+1}) = 0, \alpha_{n+1}^{\gamma} \in [0, \gamma]\}$.
  - $\mathcal{E} = \{i : f^{\gamma}(\mathbf{x}_i) < 0, \alpha_i^{\gamma} = 1, i = 1, ..., n\}\bigcup\{n+1 : f^{\gamma}(\mathbf{x}_{n+1}) < 0, \alpha_{n+1}^{\gamma} = \gamma\}$.
  - $\mathcal{C} = \{i : f^{\gamma}(\mathbf{x}_i) > 0, \ \alpha_i^{\gamma} = 0\}$, for $i = 1, ..., n+1$.

Group composition does not change as $\gamma \in [\gamma^-, \gamma^+]$ given $\alpha^{\gamma^-}, \rho^-$.

- **Decision function**:

$$f^{\gamma}(\mathbf{x}) = \frac{1}{(n+\gamma)\nu}\left(\sum_{i=1}^{n+1}(\alpha_i^{\gamma} - \alpha_i^{\gamma^-})K(\mathbf{x}, \mathbf{x}_i) - (\alpha_0^{\gamma} - \alpha_0^{\gamma^-}) + (n+\gamma^-)\nu f^{\gamma^-}(\mathbf{x})\right). \quad (3)$$

where $\alpha_0^{\gamma} = (n+\gamma)\nu\rho^{\gamma}$.

- **Determination of $\boldsymbol{\alpha}^{\gamma}$**: Let $A = \begin{bmatrix} K_{\mathcal{M}} & -\mathbf{1} \\ -\mathbf{1}^T & 0 \end{bmatrix}$, $\mathbf{c}^T = [0...0, 1]$,
  - When $f^{\gamma}(\mathbf{x}_{n+1}) > 0$ ($\mathcal{C}$) or $f^{\gamma}(\mathbf{x}_{n+1}) = 0$ ($\mathcal{M}$),

$$\begin{pmatrix} \boldsymbol{\alpha}_{\mathcal{M}}^{\gamma} \\ \alpha_0^{\gamma} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_{\mathcal{M}}^{\gamma^-} \\ \alpha_0^{\gamma^-} \end{pmatrix} + (\gamma - \gamma^-)\nu A^{-1}\mathbf{c}. \quad (4)$$

  - When $f^{\gamma}(\mathbf{x}_{n+1}) < 0$ ($\mathcal{E}$),

$$\begin{pmatrix} \boldsymbol{\alpha}_{\mathcal{M}}^{\gamma} \\ \alpha_0^{\gamma} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_{\mathcal{M}}^{\gamma^-} \\ \alpha_0^{\gamma^-} \end{pmatrix} + (\gamma - \gamma^-)A^{-1}\left((\nu - 1)\mathbf{c} - K_{\mathcal{M},n+1}\right). \quad (5)$$

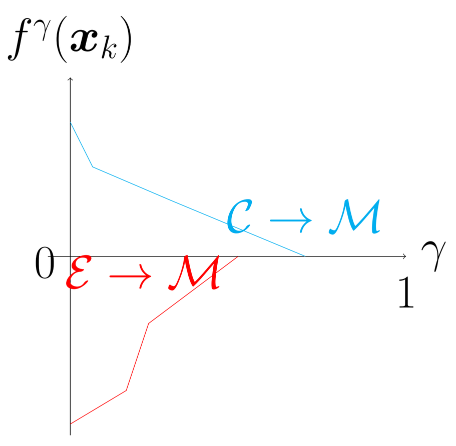
Group partition change

Let $\Delta\gamma = \gamma^+ - \gamma^-$, during the online learning, find the minimum $\Delta\gamma > 0$, such that the following events happen until $\gamma^+ = 1$.

- $\mathcal{C} \to \mathcal{M}$: $f^{\gamma}(\mathbf{x}_k) > 0$ to $f^{\gamma}(\mathbf{x}_k) = 0$, then

$$\Delta\gamma_k = \begin{cases} -\frac{(n+\gamma^-)f^{\gamma^-}(\mathbf{x}_k)}{[K_{k,\mathcal{M}}, -1]\boldsymbol{v}}, & n+1 \notin \mathcal{E}, k \in \mathcal{C}, \\ -\frac{(n+\gamma^-)f^{\gamma^-}(\mathbf{x}_k)}{[K_{k,\mathcal{M}}, -1]\boldsymbol{u} + K_{k,n+1}}, & n+1 \in \mathcal{E}, k \in \mathcal{C}, \end{cases} \quad (6)$$

- $\mathcal{E} \to \mathcal{M}$: $f^{\gamma}(\mathbf{x}_k) < 0$ to $f^{\gamma}(\mathbf{x}_k) = 0$, then $\Delta\gamma$ is the same as (6) except $f^{\gamma^-}(\mathbf{x}_k) < 0$.

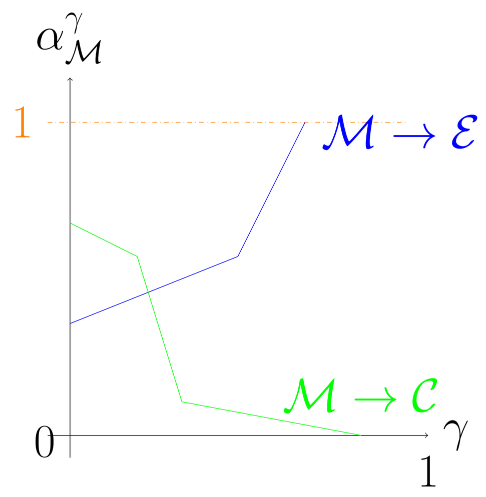


- $\mathcal{M} \to \mathcal{C}$: $\alpha_k^{\gamma} = 0$, then

$$\Delta\gamma_k = \begin{cases} -\frac{\alpha_k^{\gamma^-}}{\nu v_k}, & n+1 \notin \mathcal{E}, k \in \mathcal{M}, \\ -\frac{\alpha_k^{\gamma^-}}{u_k}, & n+1 \in \mathcal{E}, k \in \mathcal{M}. \end{cases} \quad (7)$$

- $\mathcal{M} \to \mathcal{E}$: $\alpha_k^{\gamma} = 1$ except $\alpha_{n+1}^{\gamma} = \gamma^+$, then

$$\Delta\gamma_k = \begin{cases} \frac{1-\alpha_k^{\gamma^-}}{\nu v_k}, & n+1 \in \mathcal{C} \text{ or } \mathcal{M}, k \in \mathcal{M}, k \neq n+1, \\ \frac{\alpha_{n+1}^{\gamma^-} - \gamma^-}{1 - \nu v_{n+1}}, & n+1 \in \mathcal{M}, k = n+1, \\ \frac{1-\alpha_k^{\gamma^-}}{u_k}, & n+1 \in \mathcal{E}, k \in \mathcal{M}. \end{cases} \quad (8)$$



Experiments

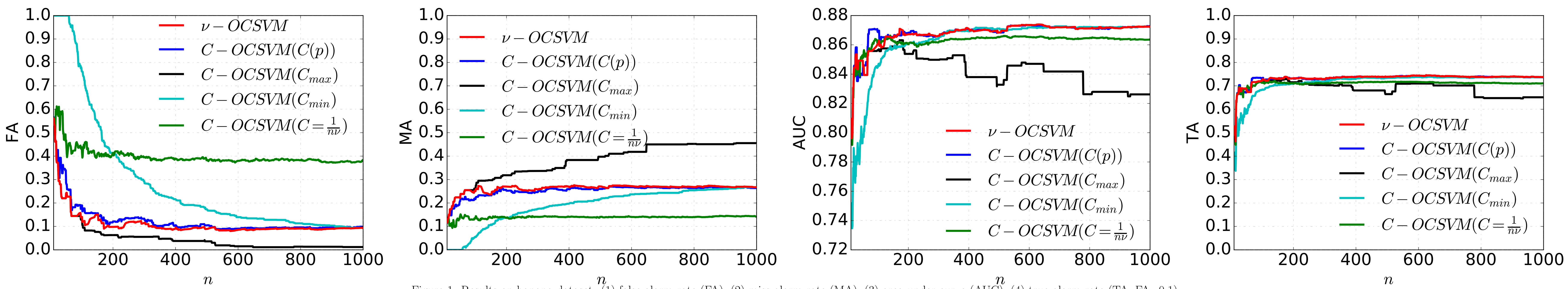


Figure 1. Results on banana dataset: (1) false alarm rate (FA), (2) miss alarm rate (MA), (3) area under curve (AUC), (4) true alarm rate (TA, FA=0.1).

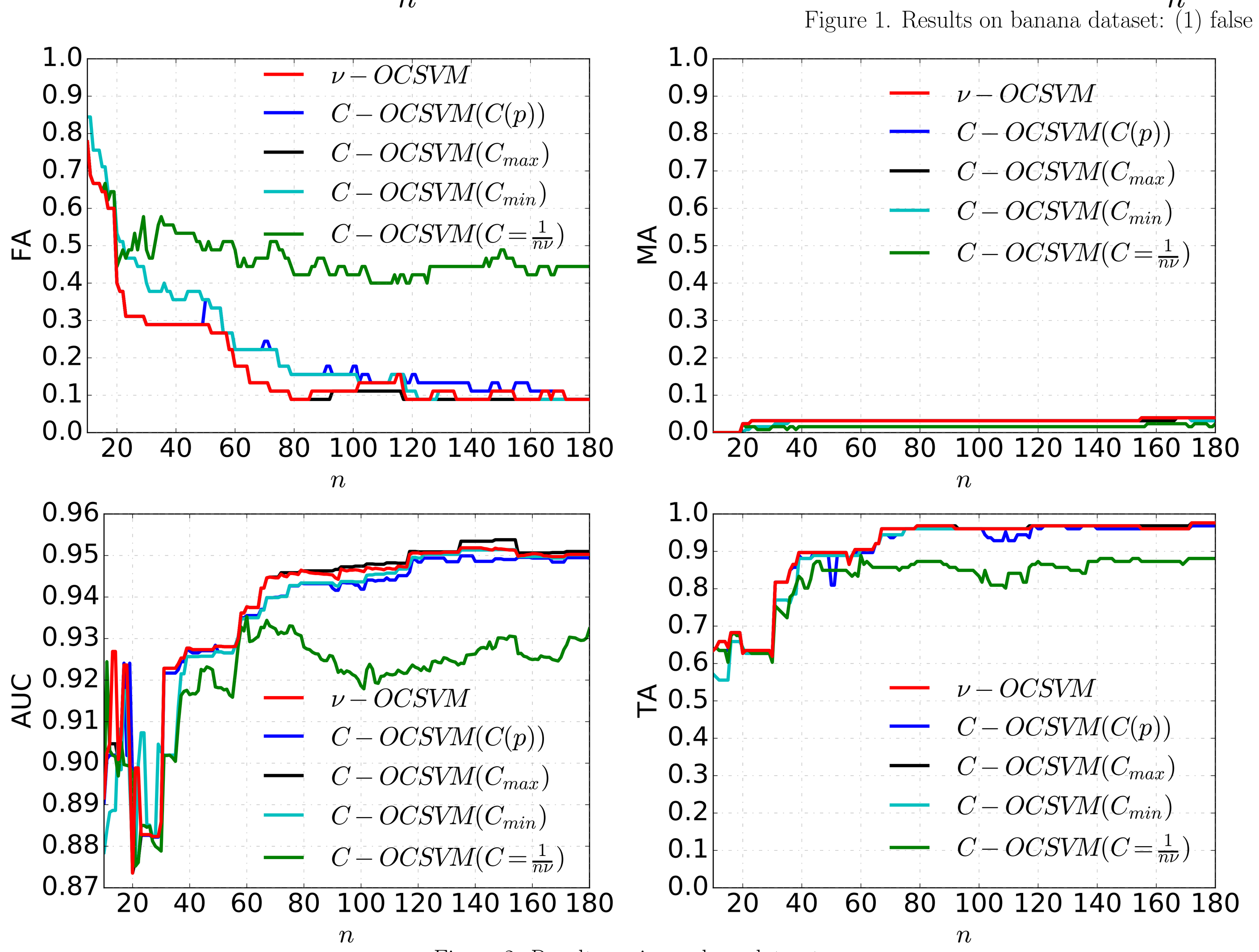


Figure 2. Results on ionosphere dataset.

Table 1. AUC and TA (FA=0.1) on other datasets.

N. samples(%)		AUC					TA (FA=0.1)				
		10%	20%	30%	50%	100%	10%	20%	30%	50%	100%
Square	ν	**0.8036**	**0.8125**	0.8131	0.8145	0.8169	**0.6228**	0.6458	**0.6537**	0.6649	**0.6699**
	$C(p)$	0.8023	0.8124	**0.8133**	**0.8151**	0.8168	0.6197	**0.6465**	0.6530	0.6643	0.6678
	$C_{\max}$	**0.8036**	**0.8125**	0.8131	0.8143	**0.8179**	**0.6228**	0.6458	**0.6537**	**0.6672**	0.6698
	$C_{\min}$	0.7868	0.8067	0.8122	**0.8151**	0.8169	0.5869	0.6356	0.6530	0.6642	**0.6699**
	$C = \frac{1}{n\nu}$	0.7656	0.7638	0.7767	0.7842	0.7930	0.5448	0.5275	0.5626	0.5750	0.6074
Spiral	ν	**0.8106**	0.8247	0.8345	0.8355	0.8448	**0.6265**	0.6515	0.6797	0.6817	0.6998
	$C(p)$	0.8102	0.8260	0.8379	**0.8388**	**0.8461**	0.6180	0.6585	0.6866	**0.6841**	**0.7020**
	$C_{\max}$	**0.8106**	0.8247	0.8345	0.8356	0.8429	**0.6265**	0.6515	0.6797	0.6818	0.6960
	$C_{\min}$	0.8005	**0.8281**	**0.8403**	0.8386	0.8448	0.5756	**0.6636**	**0.6894**	0.6826	0.6998
	$C = \frac{1}{n\nu}$	0.7661	0.7835	0.8052	0.8174	0.8241	0.5084	0.5641	0.6063	0.6382	0.6625
Digits	ν	0.9379	0.9395	0.9465	0.9435	0.9434	0.8625	0.8560	0.8649	0.8627	0.8658
	$C(p)$	**0.9463**	0.9412	0.9489	0.9457	0.9448	**0.8741**	**0.8598**	**0.8747**	0.8734	0.8665
	$C_{\max}$	0.9353	0.9258	0.9307	0.9226	0.9156	0.8515	0.8411	0.8415	0.8152	0.7838
	$C_{\min}$	0.9416	**0.9450**	0.9508	0.9524	0.9434	0.8049	0.8386	0.8716	**0.8861**	0.8658
	$C = \frac{1}{n\nu}$	0.9403	0.9436	**0.9511**	**0.9536**	**0.9532**	0.8310	0.8449	0.8625	0.8667	**0.8807**

Conclusion

- Both ν-OCSVM and C-OCSVM ($C(p)$) have stable performance as more data are added.
- Online ν-OCSVM is a good mean to target a give false alarm rate, while C-OCSVM ($C(p)$) needs two steps, the online learning and the parameter adaptation.

[1] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
[2] Pavel Laskov, Christian Gehl, Stefan Krüger, and Klaus-Robert Müller, "Incremental support vector learning: Analysis, implementation and applications," *The Journal of Machine Learning Research*, vol. 7, pp. 1909–1936, 2006.
[3] Xin Tong, Yang Feng, and Anqi Zhao, "A survey on neyman-pearson classification and suggestions for future research," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 8, no. 2, pp. 64–81, 2016.