Hyper-parameters auto-setting using regularization path for ν-SVM

Loosli Gaëlle and Canu Stéphane

Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes (LITIS)
Institut National des Sciences Appliquées (INSA) de Rouen, 1060 Avenue de l'Université, BP08
F-76801 St Etienne du Rouvray Cedex - France
Email: gaelle.loosli@insa-rouen.fr

Abstract

- Goal: auto-setting of the regularization parameter for SVM
- Tools:
  - Simple-ν-SVM solver
  - Regularization path for SVM
  - Leave-one-out validation method

ν-SVM

We consider a binary classification problem with training patterns $x_1 \dots x_m \in \mathcal{X}$ and classes $y_1 \dots y_m \in \{+1, -1\}$.

$$\begin{cases} \min_{f,b,\rho,\xi_i} \frac{m}{2}\|f\|^2 - \lambda\rho + \sum_{i=1}^{m} \xi_i \\ \text{s.t. } y_i\big(f(x_i)+b\big) \geq \rho - \xi_i & \forall i \in [1,\dots,m] \\ \text{and } \rho \geq 0 \\ \text{and } \xi_i \geq 0 & \forall i \in [1,\dots,m] \end{cases}$$

$$\begin{cases} \max_{\alpha} -\frac{1}{2}\boldsymbol{\alpha}^\top G \boldsymbol{\alpha} \\ \text{s.t. } \boldsymbol{\alpha}^\top \mathbf{1} \geq \lambda \\ \text{and } \boldsymbol{\alpha}^\top \mathbf{y} = 0 \\ \text{and } 0 \leq \alpha_i \leq 1 & \forall i \in [1,\dots,m] \end{cases}$$

where $G(i,j) = \frac{1}{m} y_i y_j k(x_i, x_j)$.

Regularization path

Our aim is to compute the ν-SVM solution for all values of ν. As shown in [1], the path is piecewise linear. It means that the support vectors set does not change between two values of ν. Hence we only need to identify when a change occurs in the sets.

Let note $g(x_i) = y_i\big(f(x_i)+b\big) - \rho$. Then we have:

$$\begin{cases} \mathcal{L} : g(x_i) < 0 \; \forall i \in \mathcal{L} & \alpha_i = 1 & \text{bounded points} \\ \mathcal{E} : g(x_i) = 0 \; \forall i \in \mathcal{E} & 0 < \alpha_i < 1 & \text{margins points} \\ \mathcal{R} : g(x_i) > 0 \; \forall i \in \mathcal{R} & \alpha_i = 0 & \text{useless points} \end{cases}$$

We identify four possible movements:

Step	$in(r)$	$out(r)$	$out(\ell)$	$in(\ell)$
t	$i \in \mathcal{R}$	$i \in \mathcal{E}$	$i \in \mathcal{E}$	$i \in \mathcal{L}$
	$g^t(x_i) > 0$	$\star\mathbf{g^t(x_i) = 0}$	$\star\mathbf{g^t(x_i) = 0}$	$g^t(x_i) < 0$
verte	$\alpha_i = 0$	$0 < \alpha_i < 1$	$0 < \alpha_i < 1$	$\alpha_i = 1$
$t+1$	$i \in \mathcal{E}$	$i \in \mathcal{R}$	$i \in \mathcal{L}$	$i \in \mathcal{E}$
	$\star\mathbf{g^{t+1}(x_i) = 0}$	$\star\mathbf{g^{t+1}(x_i) \geq 0}$	$\star\mathbf{g^{t+1}(x_i) \leq 0}$	$\star\mathbf{g^{t+1}(x_i) = 0}$
	$0 < \alpha_i < 1$	$\star\mathbf{\alpha_i = 0}$	$\star\mathbf{\alpha_i = 1}$	$0 < \alpha_i < 1$

Summary of the events. Each column stands for a particular event. In blue starred are noted the properties that are used to compute the corresponding λ^{t+1}

The Simple-ν-SVM solver

The Simple-ν-SVM solver is derived from the SimpleSVM [2]. It is an active set method, based on the repartition of the training points into three groups :

- $\mathcal{R}$ for unused points,
- $\mathcal{E}$ for support vectors placed on the margins,
- $\mathcal{L}$ for support vectors placed further.

Algorithm:

```
initialisation, ν = 1/n
while ν < 1 do
    solve linear system                              ▷ α_w
    if non admissible solution then
        project solution in the admissible set       ▷ Simple-ν-SVM step
    else if non optimal solution then
        select next point                            ▷ Simple-ν-SV step
    else
        compute the next ν                           ▷ regularization path step
    end if
end while
```

Stopping criteria

The leave-one-out error is defined as the mean error done for the removed points. We also compute a second leave-one-out estimation to have an idea of the variance of the solution:

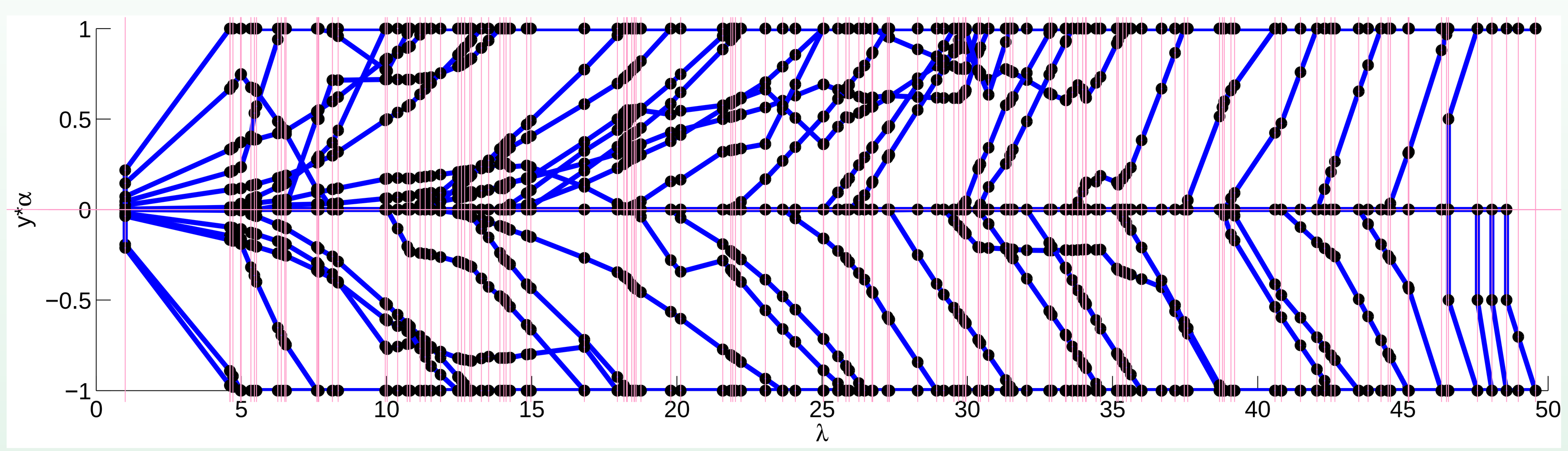
$$LOO1_{error} = \frac{1}{n}\sum_i 1 - sign(\hat{y}_i y_i) \qquad LOO2_{error} = \frac{1}{n}\sum_i max(0, \rho - \hat{y}_i y_i)$$

This second formula is very helpful to detect over-fitting. Indeed, outliers will be very penalized.

- leave-one-out error rates are estimated at each step,
- no point from $\mathcal{R}$ participate to the solution $\rightarrow$ zero error,
- need to compute the LOO errors of each point of $\mathcal{E}$ and $\mathcal{L}$ only,
- Simple-ν-SVM warm start.

Experiments

For the evaluation of our method, we have used the artificial *apple and banana* dataset.



Evaluation of the $y_i\alpha_i$ along the path for the apple and banana problem. The regularization paths can be represented via the values taken by the α coefficients during learning. Those coefficients follow piecewise linear paths. $\alpha_i y_i$ is plot, such that all negative points appear bellow zeros.

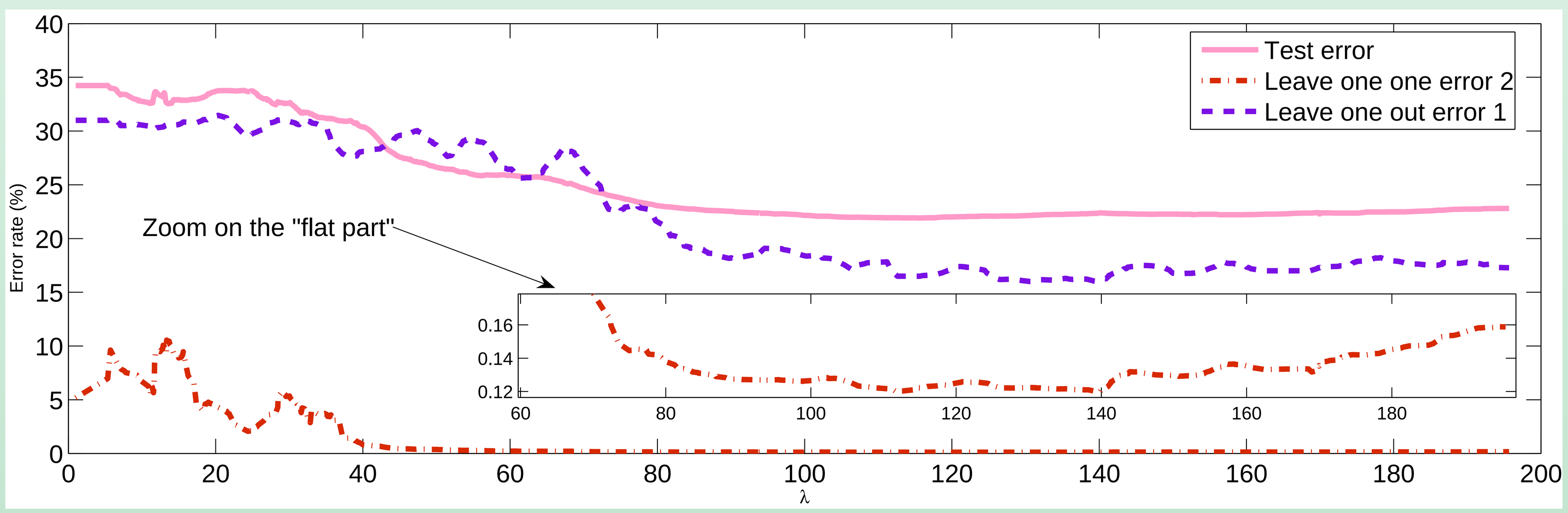


Illustration on the apple en banana dataset of the LOO error rate evolution according to λ, reported with the test error.

Acknowledgments

This work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors' views.

References

[1] Trevor Hastie, Saharon Rosset, Robert Tibshirani, and Ji Zhu. The entire regularization path for the support vector machine. *Journal of Machine Learning Research*, 5:1391–1415, 2004.

[2] Gaëlle Loosli, Stéphane Canu, SVN Vishwanathan, and Alexander J.Smola. Invariances in classification : an efficient svm implementation. *Applied Stochastic Models and Data Analysis*, 2005.

Conclusion

- ν-SVM regularization path
- stopping criteria to auto-set the regularzation parameter
- efficient ν-SVM solver
- method taking advantage of the sparcity
- allow to use regularization paths for larger databases