ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Robust outlier detection with L_0-SVDD

Meriem El Azami[1], Carole Lartizien[1] and Stéphane Canu[2] *

1- Université de Lyon, CREATIS; CNRS UMR5220; Inserm U1044;
INSA-Lyon; Univ. Lyon 1 - France

2- LITIS, INSA de Rouen, Normandie Université,
Saint-Etienne-du-Rouvray, 76801, France

Abstract. The problem of *outlier detection* consists in finding data that is not representative of the population from which it was ostensibly derived. Recently, to solve this problem, Liu et al. [1] proposed a two steps hypersphere-based approach, taking into account a confidence score pre-calculated for each input data. Defining these scores in a first step, independently from the second one, makes this approach not well-suited for large stream data. To solve these difficulties, we propose a global reformulation of the support vector data description (SVDD) problem based on the L_0 norm, well suited for outlier detection. We demonstrate that this L_0-SVDD problem can be solved using an iterative procedure providing data specific weighting terms. We show that our approach outperforms state of the art outlier detection techniques using both synthetic and clinical data.

1 Introduction

The challenging topic of *outlier detection* has gained increasing interest in supervised classification due to the difficult task of labelling samples in most application domains. In medical imaging, for instance, one major bottleneck to the development of computer assisted diagnosis (CAD) systems is building training databases where abnormal signals in the image are non-invasively associated with their corresponding histopathological ground truth. Outlier detection methods consist in learning the compact representation domain of a *normal* class, in view of predicting whether a test sample belongs to this compact description. One classical approach is the SVDD [2], which hypothesizes that the normal data belong to a hypersphere characterized by a center c and a radius R. Optimal c and R are obtained by solving the following constraint-based optimization problem, for properly chosen positive parameters C and m:

$$\begin{cases} \min\limits_{R,c,\xi} & R + C\sum\limits_{i=1}^{n}\xi_i \\ \text{with} & \|x_i - c\|^2 \leq m + R + \xi_i, \quad i = 1,\ldots,n \\ \text{and} & \xi_i \geq 0, \qquad\qquad\qquad\quad i = 1,\ldots,n \end{cases} \tag{1}$$

*This work was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). We thank Alexander Hammers from the Neurodis Foundation and Nicolas Costes from the CERMEP for providing the MRI data and for useful discussions.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

where $(x_i)_{i=1,..,n}$ are training samples from the normal class, $x_i \in \mathbb{R}^p$ and ξ_i are slack variables allowing to relax the constraints. When $m = 0$, problem (1) is the original SVDD formulation. The margin parameter m is added to deal with outlier detection, since it allows to set a confidence margin around clustered data (see fig. 1). As the formulation in (1) is based on a L_1-norm cost function, the presence of uncertain samples x_j will generate high values of ξ_j and result in a significant increase of the second term in equation 1. Fig. 1 shows an illustrative example of the impact of such an *outlier* normal point on the prediction of the hypersphere decision boundary: for small values of C, the solution provided by the m-SVDD addresses the problem of outlier detection but at the cost of a large number of support vectors and thus does not scale.

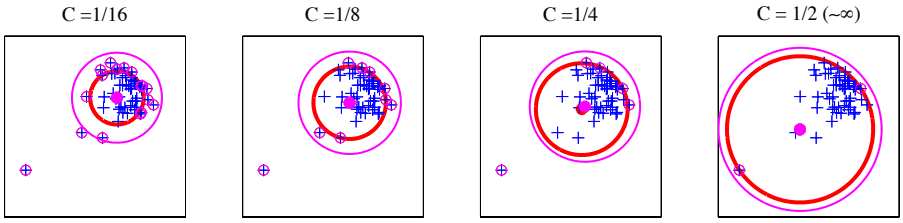

Fig. 1: Example of SVDD solutions with different C values, $m = 0$ (red) and $m = 5$ (magenta). The circled data points represent support vectors for both m.

There have been different attempts to improve the SVDD performance in the presence of uncertain data. Liu et al. [1] recently proposed to constrain the impact of the ξ_i in equation 1. Using the reference intrusion detection database, the authors showed that their approach outperforms the standard SVDD and the alternate density-induced SVDD [3]. These two approaches attempt to better control the weight of uncertain data on the cost function either by adding individual weighting terms on the ξ_i [1] or by considering data specific slack variables ξ_i [3]. In this paper, we propose to consider the L_0 cost function as an alternative to the L_1-norm in the SVDD formulation. The rational is to alleviate the influence of uncertain data by assigning them a cost of C instead of $C\xi_i$.

In the following, we first define the proposed L_0-SVDD problem and discuss the associated quadratic problem formulation. We then present results on a synthetic dataset and a real clinical application related to the detection of epileptogenic foci in MRI-based neuroimaging data. The L_0-SVDD performance is compared to that achieved by the SVDD formulation proposed by Liu.

2 Method description

We consider the problem of detecting outliers from a set of n observations of p dimensional vectors stored in X a $n \times p$ matrix. Regarding the outlier detection problem, it looks relevant to consider the zero-norm cost function [see for instance 4, for more detailed justifications]. The L_0 pseudo-norm is defined as $\|\mathrm{x}\|_0 = card\{i|x_i \neq 0\}$. The advantage of taking such a non convex cost is

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

well motivated for instance in [5] were it is shown that the resulting estimator is asymptotically unbiased. Based on this cost design for outlier detection, we propose to define the L_0-SVDD problem as follows, for a given positive C:

$$\left\{\begin{array}{ll} \min\limits_{c\in\mathbf{R}^p, R\in\mathbf{R}, \xi\in\mathbf{R}^n} & R + C\|\xi\|_0 \\ \text{with} & \|\mathrm{x}_i - c\|^2 \leq R+\xi_i \\ & \xi_i \geq 0 \qquad i = 1, n \end{array}\right.$$

Unfortunately the L_0 pseudo-norm is non differentiable, combinatorially hard, and does not lead to an effective algorithmic approach. In fact, to obtain an efficient technique for solving this problem, two key insights are needed. The first key step is, following [6], to replace the L_0 pseudo-norm by its logarithmic approximation leading to the following problem, for given parameters C and γ [1]:

$$\left\{\begin{array}{ll} \min\limits_{c\in\mathbf{R}^p, R\in\mathbf{R}, \xi\in\mathbf{R}^n} & R + C\sum\limits_{i=1}^{n} \log(\gamma + \xi_i) \\ \text{with} & \|\mathrm{x}_i - c\|^2 \leq R+\xi_i \\ & \xi_i \geq 0 \qquad i = 1, n\,. \end{array}\right.$$

This problem remains non convex. Our second key idea is to solve this problem by using an iterative procedure solving at each iteration a convex QP problem, resulting from the decomposition of the non-convex function as a difference of convex functions (DC) [7][8]. In our case this decomposition is of the form:

$$\log(\gamma + t) = f(t) - g(t) \qquad \text{with } f(t) = t \quad \text{and} \quad g(t) = t - \log(\gamma + t),$$

both functions f and g being convex. The DC framework consists in minimizing iteratively (R plus a sum of) the following convex term:

$$f(\xi) - g'(\xi)\xi = \xi - \left(1 - \frac{1}{\gamma + \xi^{\text{old}}}\right)\xi = \frac{\xi}{\gamma + \xi^{\text{old}}},$$

where ξ_i^{old} denotes the solution at the previous iteration.

The DC idea applied to our L_0-SVDD approximation consists in building a sequence of solutions of the following adaptive SVDD:

$$\left\{\begin{array}{ll} \min\limits_{c\in\mathbf{R}^p, R\in\mathbf{R}, \xi\in\mathbf{R}^n} & R + C\sum\limits_{i=1}^{n} w_i\xi_i \\ \text{with} & \|\mathrm{x}_i - c\|^2 \leq R+\xi_i \\ & \xi_i \geq 0 \qquad i = 1, n \end{array}\right. \qquad \text{with} \qquad w_i = \frac{1}{\gamma + \xi_i^{\text{old}}}.$$

Stationary conditions of the KKT give: $c = \sum_{i=1}^n \alpha_i x_i$ and $\sum_{i=1}^n \alpha_i = 1$, where α_i is the Lagrange multiplier associated with the inequality constraint $\|\mathrm{x}_i - c\|^2 \leq R + \xi_i$. The dual of this problem is [see for instance 1]:

$$\left\{\begin{array}{ll} \min\limits_{\alpha\in\mathbf{R}^n} & \alpha^\top XX^\top\alpha - \alpha^\top diag(XX^\top) \\ \text{with} & \sum_{i=1}^n \alpha_i = 1 \qquad\qquad 0 \leq \alpha_i \leq Cw_i \qquad i = 1, n \end{array}\right. \tag{2}$$

[1] Note that setting γ to one would be enough to perform the L_0 norm approximation.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

This is a classical QP with box constraints that can be solved efficiently using for instance an effective active set solver[2] [9]. Note that if λ denotes the Lagrange multiplier associated with the equality constraint $\sum_{i=1}^{n} \alpha_i = 1$, we can see that $R = \lambda + c^\top c$ by calculating the dual of (2), that is the bidual.

Put all together, this leads to the following algorithm 1.

Data: X, y, C , γ
Result: R , c, ξ , α
$w_i = 1; \quad i = 1, n;$
while *not converged* **do**
  $(\alpha, \lambda) \leftarrow$ solve_QP(X, C, w) % solve problem (2)
  $c \leftarrow X^\top \alpha;$
  $R \leftarrow \lambda + c^\top c;$
  $\xi_i \leftarrow \max(0, \|\mathrm{x}_i - c\|^2 - R) \quad i = 1, n;$
  $w_i \leftarrow 1/(\gamma + \xi_i) \quad i = 1, n;$
end

Algorithm 1: L_0-SVDD for the linear kernel

The kernelization of this algorithm is straight forward by using the kernel trick and associated representer theorem (not included here due to lack of space).

3 Experiments

3.1 Synthetic data results

Learning dataset: We generated learning examples for the normal class by drawing $n = 25$ pseudo-random normally distributed samples of dimension $p = 2$. The mean was set to 1 for the first dimension and to 2 for the second dimension. To simulate the presence of outliers in the learning dataset, we added examples that fall outside the normal class distribution range (see Fig. 2 left). The hyper-parameters were fixed as follows: $\gamma = 1$, $nb\_iter = 3$ and $C = 0.4$.

Prediction results: Fig. 2 (left), shows that 1) the L_0-SVDD decision boundary best fits the learning data 2) the Liu-SVDD sphere center is still influenced by the outlier example. Note that the standard SVDD cannot be tuned to match the results (cf. fig. 1).

3.2 Realistic data results

We compared the performance of the Liu-SVDD and the L_0-SVDD methods in detecting abnormalities in brain magnetic resonance images (MRIs) of patients suffering from intractable epilepsy on a voxelwise basis.

Learning database: In [10], we recently showed that two parametric maps extracted from the MR scans, namely the junction and extension maps help discriminating between controls and patients suffering from intractable epilepsy.

[2]available at `asi.insa-rouen.fr/enseignants/~arakoto/toolbox`

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

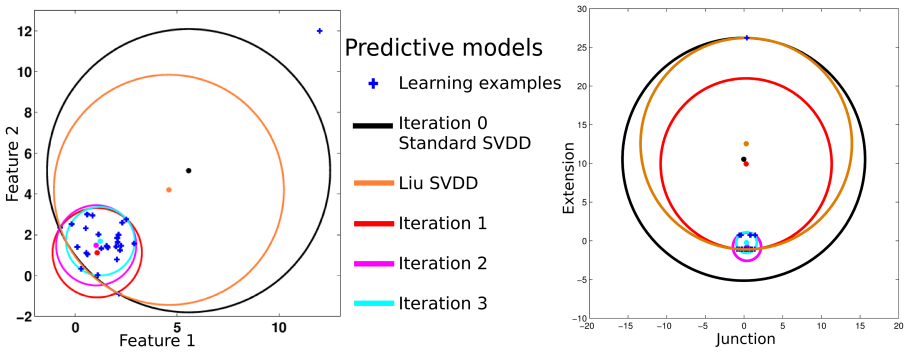

Fig. 2: Predictive models obtained using synthetic and realistic learning examples with an outlier.

For each voxel k from the brain MR scan, two classifiers, L_0-SVDD and Liu-SVDD were trained using the matrix $X^k \in M_{n,p}(\mathbb{R})$, $n = 29$ and $p = 2$. The value $C = 0.6$ was obtained from a leave one voxel out procedure for C values in $[0.1 : 0.1 : 1]$. The optimization was only performed on a subset of 4000 voxels, randomly selected from the whole brain (1.5 million voxels) to reduce computational cost. The analysis of the training data in M^k indicated that some clusters of voxels contained uncertain data resulting from artefacts in the original scan or from image processing issues. Fig. 2 (right) illustrates the presence of such uncertain data in a voxel belonging to the cluster highlighted in green in Fig. 3 (left). The distribution of the two features (junction and extension) was computed over the 29 control subjects from the learning database.

Test data: In the MRI of a control subject, at the known location of uncertain 'normal' data, we simulated an heterotopy like lesion (Fig. 3 (left)), which is an abnormal extension of the grey matter into the white matter. We locally changed the grey level values of the voxels within the white matter in the original MRI, to make them correspond to the grey matter distribution.

Prediction results: Fig. 3 shows that the L_0-SVDD classifier detected most of the lesion (DICE of 80% for the example slice in Fig. 3 and 53% for the whole lesion) while the Liu-SVDD classifier failed in retrieving the lesion (DICE of 0%). The lesion detected by L_0-SVDD is bigger than the real lesion due to smoothing in the pre-processing steps.

4 Conclusion

The conducted experiments on synthetic and realistic clinical data show that, unlike state of the art methods, the L_0-SVDD approach successfully suppresses the effect of uncertain data on the predicted decision boundary. Future work will focus on performing a quantitative analysis as in [1], evaluating the impact of the parameter γ and analysing convergence properties of the proposed algorithm.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

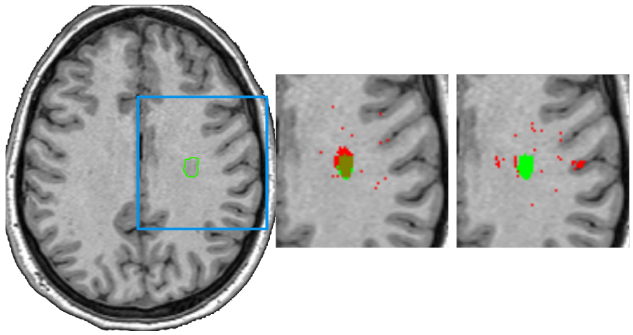

Fig. 3: Transverse MR slice of the test subject showing: the simulated lesion highlighted in green in all images, L_0-SVDD (*middle, in red*) and Liu-SVDD (*right, in red*) classification results.

References

[1] Bo Liu, Yanshan Xiao, Longbing Cao, Zhifeng Hao, and Feiqi Deng. SVDD-based outlier detection on uncertain data. *Knowledge and information systems*, 34(3):597–618, 2013.

[2] David M. J. Tax and Robert P.W Duin. Support vector data description. *Mach. Learn.*, 54(1):45–66, 2004.

[3] KiYoung Lee, Dae-Won Kim, K.H. Lee, and Doheon Lee. Density-induced support vector data description. *IEEE Trans. on Neural Netw.*, 18(1):284–289, 2007.

[4] Pedro A Forero, Vassilis Kekatos, and Georgios B Giannakis. Robust clustering using outlier-sparsity regularization. *IEEE Trans. Signal Process.*, 60(8):4163–4177, 2012.

[5] Anestis Antoniadis, Irène Gijbels, and Mila Nikolova. Penalized likelihood regression for generalized linear models with non-quadratic penalties. *Annals of the Institute of Statistical Mathematics*, 63(3):585–615, 2011.

[6] Jason Weston, André Elisseeff, Bernhard Schölkopf, and Mike Tipping. Use of the zero norm with linear models and kernel methods. *J. Mach. Learn. Res.*, 3:1439–1461, 2003.

[7] Gilles Gasso, Alain Rakotomamonjy, and Stéphane Canu. Recovering sparse signals with a certain family of nonconvex penalties and DC programming. *IEEE Trans. Signal Process.*, 57(12):4686–4698, 2009.

[8] LeThiHoai An and PhamDinh Tao. The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Annals of Operations Research*, 133(1-4):23–46, 2005.

[9] Gaëlle Loosli and Stéphane Canu. Comments on the core vector machines: Fast SVM training on very large data sets. *J. Mach. Learn. Res.*, 8:291–301, 2007.

[10] Meriem El Azami, Alexander Hammers, Nicolas Costes, and Carole Lartizien. Computer aided diagnosis of intractable epilepsy with MRI imaging based on textural information. In *International Workshop on Pattern Recognition in Neuroimaging (PRNI), 2013*, pages 90–93, June 2013.