Support Vector Machine with spatial regularization for pixel classification

R. Flamary
Lagrange Lab., Université de Nice Sophia-Antipolis,
CNRS, Observatoire de la Côte d'Azur, Nice, France
remi.flamary@unice.fr

A. Rakotomamonjy
LITIS Lab.
Université de Rouen, France
alain.rakoto@insa-rouen.fr

Abstract: We propose in this work to regularize the output of a svm classifier on pixels in order to promote smoothness in the predicted image. The learning problem can be cast as a semi-supervised SVM with a particular structure encoding pixel neighborhood in the regularization graph. We provide several optimization schemes in order to solve the problem for linear SVM with ℓ_2 or ℓ_1 regularization and show the interest of the approach on an image classification example with very few labeled pixels.

Keywords: Support Vector Machine, Large scale learning, semi-supervised learning

1 Introduction

Pixel classification is the problem of assigning a class to every pixel in an image. This is a classical problem with several applications in medical imaging or in geoscience remote sensing where it is denoted as image classification [2, 6]. A common approach for solving this problem is to use discriminative machine learning techniques and to treat pixels as independent vectors. In order to take into account the spatial prior over the pixels, several approaches have been proposed. One example is to include spatial features or kernels to the pixel representation such as filter output images [6]. Another approach is to use a post-processing on the output of the classifier, for instance by using a Markov Random Field to include spatial information [8]. While the post-processing approach can integrate high order relations between pixels, it is also more computationally intensive.

Another challenge of pixel classification is the dataset itself. The number of pixel N increases quadratically with the size of the image and the number n of labeled pixels is usually small. This suggest the use of semi-supervised learning methods [2] which have led to dramatic performance improvement when the number of labeled pixels is small. Note that in their works, the large number of pixels are handled using low-rank kernel approximations, leading to the learning of a linear SVM on a small number d of nonlinear features.

In this paper, we focus on linear SVM applied on $d \ll N$ features (potentially nonlinear) extracted form the data. We want not only to use unlabeled pixels in the learning problem but also to promote spatial smoothness on the output of the prediction function, thus using unlabeled pixels. We propose to this end to regularize the SVM output using a term that encodes the spatial neighborhood of the pixels as seen in [5]. This approach is a particular case of manifold-based regularized semi-supervised learning. We discuss in the following how to solve the learning problem for different regularizations on the linear SVM. Finally, numerical experiments are performed in order to show the interest of the approach on a difficult pixel classification problem.

2 SVM with spatial regularization

The dataset consists in a full image of N pixels with d features per pixel (possibly hyperspectral spectrum or other features). These pixels $\mathbf{x}$ are stored in the matrix $\mathbf{X} \in \mathbb{R}^{N\times d}$. Only $n < N$ of these pixels with indexes $i \in \mathcal{L}$ are labeled with $y_i \in \{-1, +1\}$. We want to learn a prediction function $f(\cdot)$ of the form

$$f(\mathbf{x}) = \sum_i w_i x_i + b = \mathbf{w}^\top \mathbf{x} + b \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^d$ is the normal vector to the separating hyperplane and $b \in \mathbb{R}$ is a bias term.

2.1 Learning problem

We propose to learn the prediction function with the following optimization problem:

$$\min_f \sum_{i\in\mathcal{L}} H(y_i, f(\mathbf{x}_i)) + \lambda_s \sum_{i,j} W_{i,j}(f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 + \lambda_r \Omega(f) \tag{2}$$

where $H(y, f(\mathbf{x})) = \max(0, 1 - yf(\mathbf{x}))^2$ is the squared hinge loss, $\mathbf{W} \in \mathbb{R}^{N\times N}$ is a symmetric matrix of general term $W_{i,j}$ that encodes the similarity between pixel i and j, λ_s and λ_r are regularization parameters and $\Omega(\cdot)$ is the SVM regularization term.

This problem is a classical semi-supervised learning problem. If the similarity matrix $\mathbf{W}$ were chosen to be a Gaussian kernel matrix, the problem would boil down

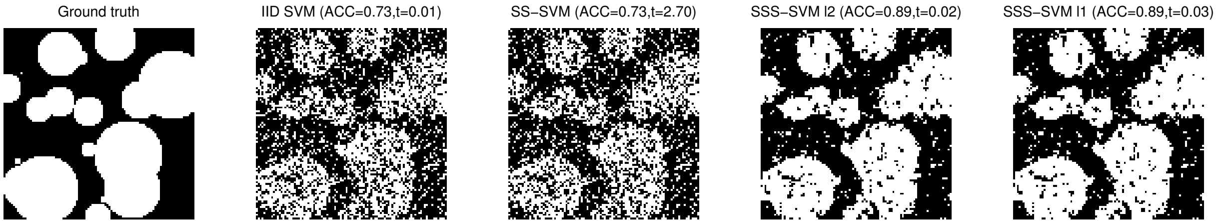

Fig. 1: Ground truth labels, accuracy ACC, training time t in second and decision maps for IID SVM, SS-SVM and SSS-SVM with ℓ_2 and ℓ_1 regularization. The regularization parameters of each methods are selected in order to maximize test accuracy.

to a Laplacian SVM [2]. But it requires the computation of a $\mathcal{O}(N^2)$ kernel matrix. In our case, we want to promote smoothness in the output on the prediction function *i.e.* we want neighbor pixels to have similar prediction score. To this end, we propose a $\mathbf{W}$ matrix such that $W_{i,j} = 0$ everywhere except when pixels $\mathbf{x}_i$ and $\mathbf{x}_j$ are spatial neighbors ($W_{i,j} = 1$). Note that this regularization is similar to a total variation regularization but with a quadratic penalty term. Moreover the Laplacian regularization term can be computed with a complexity $\mathcal{O}(N)$ which is essential to large scale learning.

2.2 Optimization algorithm

Problem (2) can be reformulated in the linear case as

$$\min_{\mathbf{w},b} \sum_{i\in\mathcal{L}} H(y_i, \mathbf{w}^\top \mathbf{x}_i + b) + \lambda_s \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} + \lambda_r \Omega(\mathbf{w}) \quad (3)$$

where $\mathbf{\Sigma} = \mathbf{X}^\top(\mathbf{D} - \mathbf{W})\mathbf{X}$ with $\mathbf{D}$ the diagonal matrix such that $D_{i,i} = \sum_{j=1}^{N} W_{i,j}$.

When $\Omega(\mathbf{w}) = \|\mathbf{w}\|_2^2$, the problem is a classical ℓ_2 SVM with a metric regularization ($\widetilde{\mathbf{\Sigma}} = \mathbf{\Sigma} + \lambda_r/\lambda_s \mathbf{I}$). One approach suggested by [7] and [5] is to perform a change of variable $\tilde{\mathbf{w}} = \mathbf{\Sigma}^{1/2}\mathbf{w}$ and $\tilde{\mathbf{x}} = \mathbf{\Sigma}^{-1/2}\mathbf{x}$. The resulting problem can be solved with a classical linear SVM solver such as the one proposed by [3].

When $\Omega(\cdot)$ is a more complex regularization term such as the ℓ_1 norm, we propose to use a proximal splitting algorithm such as ADMM to solve the problem [1]. This approach allows us to use iteratively the efficient solver discussed above while integrating prior information to the problem through regularization.

3 Numerical experiments

Numerical experiments are performed on a simulated image of size (100×100). The simulated image that can be seen in Fig. 1 is generated as follows: i) the ground truth image is obtained by generating random circles in the image that are set to $+1$ (-1 for the background), ii) 10 discriminant features are generated by applying Gaussian noise to the ground truth image ($\sigma = 5$), iii) the previous images are filtered by a 3×3 average filter and a 3×3 median filter resulting in 20 additional features. iv) 10 images containing only Gaussian noise are added to obtain 40 features.

In order to demonstrate the interest of our approach we randomly select 10 labeled samples from each classes and we learn an independent SVM (IID-SVM), a semi-supervised Laplacian SVM (SS-SVM), and our proposed approach, the spatially regularized semi-supervised SVM (SSS-SVM) for both ℓ_2 and ℓ_1 regularization. Results show that smooth classification and prediction maps are enforced leading to an important improvement in recognition performances (see Fig. 1).

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[2] G. Camps-Valls, T. Bandos Marsheva, and D. Zhou. Semi-supervised graph-based hyperspectral image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(10):3044–3054, 2007.

[3] O. Chapelle. Training a support vector machine in the primal. *Neural Comput.*, 19(5):1155–1178, 2007.

[4] M. Dundar, B. Krishnapuram, J. Bi, and R.B. Rao. Learning classifiers when the training data is not iid. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2007.

[5] M. Dundar, J. Theiler, and S. Perkins. Incorporating spatial contiguity into the design of a support vector machine classifier. In *IGARSS*, pages 364–367. IEEE, 2006.

[6] M. Fauvel, J. Chanussot, and J. A. Benediktsson. A spatial-spectral kernel-based approach for the classification of remote-sensing images. *Pattern Recogn.*, 45:381–392, January 2012.

[7] V. Sindhwani, P. Niyogi, M. Belkin, and S. Keerthi. Linear manifold regularization for large scale semi-supervised learning. In *Proc. of the 22nd ICML Workshop on Learning with Partially Classified Training Data*, 2005.

[8] Y. Tarabalka, M. Fauvel, J. Chanussot, and J.A. Benediktsson. Svm-and mrf-based method for accurate classification of hyperspectral images. *Geoscience and Remote Sensing Letters, IEEE*, 7(4):736–740, 2010.