Max-Margin Semi-NMF

Vijay Kumar B G
vijay.kumar@elec.qmul.ac.uk
Irene Kotsia
irene.kotsia@eecs.qmul.ac.uk
Ioannis Patras
i.patras@eecs.qmul.ac.uk

Multimedia and Vision Research Group
Queen Mary, University of London
London, UK

The Non-Negative Matrix Factorization (NMF) algorithm is one of the most popular Machine Learning techniques, being widely used for many computer vision applications. NMF aims at decomposing the data matrix into a product of a non-negative basis matrix with a non-negative coefficients matrix. Regarding the discriminative power of the features extracted using NMF, only a few approaches have been proposed. More precisely, in [2] the authors introduced discriminative constraints in order to extract bases that correspond to discriminative facial regions for the problem of face recognition. The proposed Discriminant NMF (DNMF) [2] resulted in bases corresponding to salient facial features, such as eyes, mouth etc. In [1] projected gradients were used in DNMF (PGDNMF) for facial expression and face recognition. In both of the above mentioned approaches the discriminant constraints were introduced in the cost function to yield more discriminative bases. However, the introduced constraints were taylored for a rather simplistic LDA-based classifier.

In the proposed approach we choose the projections in such a way that the discriminative ability of an SVM classifier is maximized, therefore ensuring a higher classification performance. More specifically, we introduce soft max-margin constraints to the objective function of NMF to obtain a basis matrix that maximizes the classification margin using the features that are extracted using those bases. The optimization is performed with respect to the unknown bases, the projection coefficients and the parameters of the separating hyperplane and is solved in an iterative manner, where at each iteration we solve only for one of them while keeping the others fixed. The resulting sub-optimization problems are either instances of Quadratic programming with linear inequality constraints or classical SVM-type problems.

Let $\{\mathbf{x}_i, y_i\}_{i=1}^{L}$ denote a set of data vectors and their corresponding labels, where $\mathbf{x}_i \in \mathbb{R}^m$, $y_i \in \{-1,1\}$. The objective is to determine a set of basis vectors that can be used to extract features that are optimal under a max-margin classification criterion. The optimization problem for the proposed criterion is given by

$$\underset{\mathbf{G},\mathbf{H},\mathbf{w},b,\xi_i}{\operatorname{argmin}} \ \lambda\|\mathbf{X}-\mathbf{G}\mathbf{H}\|_F^2 + \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{L}\xi_i \tag{1}$$
$$\text{s.t. } y_i(\mathbf{w}^T\mathbf{G}^T\mathbf{x}_i + b) \geqslant 1-\xi_i$$
$$\xi_i > 0, \ \ 1 \le i \le L, \ \mathbf{H} \ge 0$$

where $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{L}$, $\mathbf{G}$, $\mathbf{H}$ the decomposition matrices, λ and C are positive constants and $\mathbf{w}, b, \xi_i$ are the classifier parameters. The first term in the above optimization problem corresponds to the NMF reconstruction error while the remaining terms correspond to the maximum margin classifier. The above formulation aims at maximizing the margin of the support vectors while at the same time minimizing the reconstruction and misclassification error. The classifier is trained on the projected data points $\mathbf{G}^T\mathbf{x}$, obtaining in this way the hyperplane parameter $\mathbf{w}$ and b. We iteratively solve for one of the terms $\mathbf{G}$, $\mathbf{H}$ and $\mathbf{w}, b, \xi_i$ by keeping the remaining parameters fixed

We first solve for $\mathbf{G}$ by keeping $\mathbf{H}, \mathbf{w}$ and b fixed. Since $\mathbf{w}$ is fixed, the optimization problem in Eqn. 1 is simplified as

$$\underset{\mathbf{G},\xi_i}{\operatorname{argmin}} \ \ \lambda\|\mathbf{X}-\mathbf{G}\mathbf{H}\|_F^2 + C\sum_{i=1}^{L}\xi_i \tag{2}$$
$$\text{s.t. } y_i(\mathbf{w}^T\mathbf{G}^T\mathbf{x}_i + b) \geqslant 1-\xi_i$$
$$\xi_i > 0, \ \ 1 \le i \le L$$

The above formulation is a weighted combination of the reconstruction error (1st term) and soft constraints/penalizations for the examples that do not maintain the appropriate distance (margin) from the separating hyperplane (2nd term). In the next step, we keep the basis $\mathbf{G}$ and weight matrix $\mathbf{H}$ fixed and determine a hyperplane that maximizes the margin of the classifier. The features are obtained by projecting the data points onto the updated basis matrix.Since $\mathbf{G}$ and $\mathbf{H}$ are fixed, the optimization problem in Eqn. 1 is simplified to that of a classical SVM which is solved using a off-the-shelf SVM classifier. Finally we solve for the weight matrix $\mathbf{H}$ by keeping $\mathbf{G}$, $\mathbf{w}$ and b fixed.

During testing, the input test vector $\mathbf{x}_{test}$ is projected onto the basis matrix to obtain the feature vector, $\mathbf{f}_{test} = \mathbf{G}^T\mathbf{x}_{test}$. The feature vector is applied at the max-margin classifier which predicts the class $\hat{\mathbf{y}}_{test} = sign(\mathbf{w}^T\mathbf{f}_{test} + b)$ where $\mathbf{w}, b, \mathbf{G}$ are computed during training.

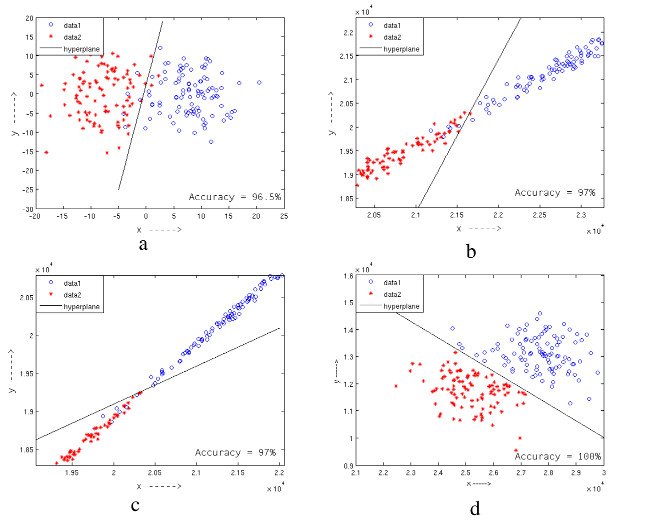

Figure 1: The projections and the SVM separating hyperplane using (*a*) PCA (*b*) Semi-NMF bases. (*c*) Max-margin NMF bases (1st iteration) and (*d*) Max-margin NMF bases (6^{th} iteration) respectively.

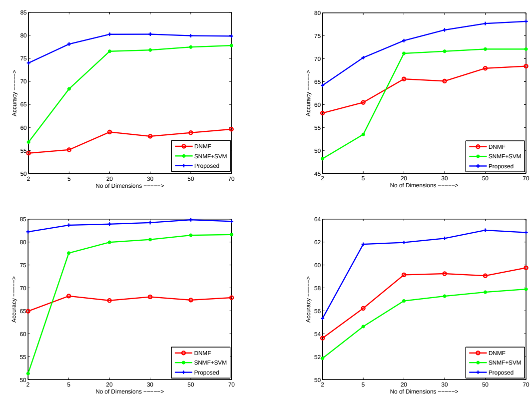

Figure 2: Comparison of the performance of the proposed algorithm with DNMF [2], Semi-NMF + SVM on different categories of Mediamill dataset. The graph shows accuracies computed at different number of bases *k*.

[1] I. Kotsia, S. Zafeiriou, and I. Pitas. A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems. *IEEE Transactions on Information Forensics and Security*, 2(3):588–595, 2007.

[2] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas. Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification. *IEEE Transactions on Neural Networks*, 17 (3):683–695, 2006.