Group Sparse Non-negative Matrix Factorization for Multi-Manifold Learning

Xiangyang Liu [1,2]
liuxy@sjtu.edu.cn
Hongtao Lu [1]
htlu@sjtu.edu.cn
Hua Gu [2]
guhuasy@hhu.edu.cn

[1] Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, China

[2] College of Science
Hohai University
Nanjing, China

Many observable data sets can be modeled by a mixture of manifolds [2]. For example, for handwritten digits, each digit forms its own manifold in the feature space. For human faces, the face images of the same person under different conditions lie on the same manifold and different persons are associated with different manifolds. Many works focused on the case of hybrid linear modeling, i.e., one linear model for each homogeneous subset of data [1, 6]. Recently, many group sparsity regularization methods have been presented such as group lasso [5] and so on.

In this paper, we propose a novel algorithm to learn multiple linear manifolds based on group sparsity and non-negative matrix factorization. Via the group sparsity constraint imposed on the column vectors of the coefficient matrix, we obtain multiple linear manifolds each of them belongs to a particular class. We adopt the l_1/l_2 regularizer for the objective function to yield group sparsity. For a test image, we represent it as a linear combination of the learned linear manifolds, and then the representation is naturally group sparse: only the coefficients corresponding to the same class are nonzero. The proposed algorithm achieves good recognition rates on face images with varying illuminations and expressions.

Given a data matrix $X = [x_1, x_2, \ldots, x_n] \in \Re_+^{m\times n}$, Our GSNMF also aims to find two non-negative matrices $W \in \Re_+^{m\times r}$ and $H \in \Re_+^{r\times n}$. Assuming the number of manifolds is K, and the dimension of each manifold is p. Thus $r = K \times p$ in our scheme. For the j-th column vector h_j of H, it can be divide into K groups $\mathcal{G}_k, k = 1, ..., K$, and each group has p coefficients. Given a grouping $\mathcal{G}$ of the column vector h_j of H, the kth group norm $\|\mathcal{G}_k\|_2$ is given by

$$\|\mathcal{G}_k\|_2 = \left(\sum_{\alpha\in\mathcal{G}_k} H_{\alpha j}^2\right)^{\frac{1}{2}}, \tag{1}$$

where $\|\cdot\|_2$ is l_2 norm. Then the group sparsity for the column $h_j (j = 1, ..., n)$ of coefficient matrix H is defined by

$$\|h_j\|_1^{\mathcal{G}} = \sum_k \|\mathcal{G}_k\|_2 = \sum_k \left(\sum_{\alpha\in\mathcal{G}_k} H_{\alpha j}^2\right)^{\frac{1}{2}}, \tag{2}$$

where $\|\cdot\|_1$ is l_1 norm. $\|h_j\|_1^{\mathcal{G}}$ is the l_1/l_2 regularizer of the vector h_j.

After the l_1/l_2 regularizer on the column vectors of the coefficient matrix H, the corresponding basis matrix W is expected to be composed of multiple linear manifolds each of which belongs to a particular class. GSNMF minimizes the distance objective function combining the l_1/l_2 regularizer as follows

$$O = \|X - WH\|_F^2 + \lambda \sum_{j=1}^{n} \|h_j\|_1^{\mathcal{G}}, \quad s.t.\ W \geq 0, H \geq 0, \tag{3}$$

where the regularization parameter λ controls the smoothness of the new representation.

The updating rules of Eq. (3) can easily be achieved as follows

$$H_{tj} \leftarrow H_{tj} \frac{(W^T X)_{tj}}{(W^T WH)_{tj} + \frac{\lambda H_{tj}}{2\sqrt{\sum_{t,\alpha\in\mathcal{G}_k} H_{\alpha j}^2}}}, \tag{4}$$

$$W_{it} \leftarrow W_{it} \frac{(XH^T)_{it}}{(WHH^T)_{it}}. \tag{5}$$

The convergence of the updating formula can be proved using the similarity method in [4] with the auxiliary function.

In theory, NMF unavoidably converges to local minima [4], and the solution to Eq. (3) is not unique under such constraints. We will give a proper initialization in order to improve the performance of GSNMF either for consideration of computational complexity or for the learning of multiple linear manifolds. The initialization is given as follows. For each class, we learn a basis matrix $W_i \in \Re^{m\times p}, (i = 1, ..., K)$ from the training data by NMF. Then W can be initialized by $W = [W_1, W_2, ..., W_K]$, where K is the number of manifolds. At the same time, in order to enforce the multiple linear manifolds learning, the corresponding coefficient matrix is initialized by

$$H_{ij} = \begin{cases} \delta/p & \lceil i/p \rceil = \lceil j/q \rceil \\ (1-\delta)/(r-p) & otherwise \end{cases}, \tag{6}$$

where $\lceil x \rceil$ is the ceiling function which rounds x to the nearest integers towards infinity, p is the dimension of each manifold, q is the number of the training data for each class, $r = K \times p$ and $\delta > 0$ is close to 1.

After the matrix factorization, we gain a nonnegative basis matrix W and the coefficient matrix H. Ideally, each column in the matrix W should represent a human subject. We visualize these basis learned by PCA, NMF, SNMF and GSNMF in Figure 1. We can easily see that GSNMF learns private representation for each class varying in facial expression, eye wear, pose and lighting (Each column in Figure 1 (c) belongs to a manifold). The main reason is that group sparsity constraint imposed on the coefficient matrix H leads to multiple linear manifolds in W. For face recognition on the ORL, Yale and Extended Yale B databases, our algorithm GSNMF outperforms Principal Component Analysis (PCA), the original NMF [4] and the Sparse NMF (SNMF) [3].

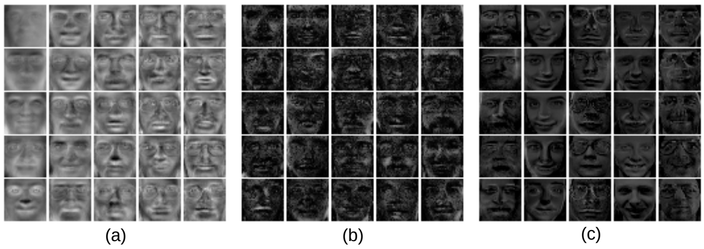

Figure 1: Basis vectors learned from the face images in ORL database. (a) shows the first 25 Eigenfaces, (b) shows 25 NMF faces, and (c) shows 25 GSNMF faces.

[1] G. Chen and G. Lerman. Spectral curvature clustering (SCC). *International Journal of Computer Vision*, 81(3):317–330, 2009. ISSN 0920-5691.

[2] Ronald A. DeVore, Tamara G. Kolda, Gilad Lerman, and Guillermo R. Sapiro. Hot topics workshop: Multi-manifold data modeling and applications, 2008. URL `http://www.ima.umn.edu/2008-2009/SW10.27-30.08/`.

[3] P.O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004. ISSN 1532-4435.

[4] D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999. ISSN 0028-0836.

[5] L. Meier, S. van de Geer, and P. Buhlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society Series B*, 70(1):53–71, 2008. ISSN 1369-7412.

[6] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1945–1959, 2005.