# Neighborhood Preserving Nonnegative Matrix Factorization

Quanquan Gu
gqq03@mails.tsinghua.edu.cn
Jie Zhou
jzhou@tsinghua.edu.cn

State Key Laboratory on Intelligent Technology and Systems
Tsinghua National Laboratory for Information Science and Technology (TNList)
Department of Automation, Tsinghua University, Beijing 100084, China

*Nonnegative Matrix Factorization* (NMF) [2] has been widely used in computer vision and pattern recognition. It aims to find two nonnegative matrices whose product can well approximate the nonnegative data matrix, which naturally leads to parts-based and non-subtractive representation. Recent years, many variants of NMF have been proposed. [3] proposed a local NMF (LNMF) which imposes a spatially localized constraint on the bases. All the methods mentioned above are unsupervised, while [5] proposed a discriminative NMF (DNMF), which adds an additional constraint seeking to maximize the between-class scatter and minimize the within-class scatter in the subspace spanned by the bases.

Recent studies have shown that many real world data are actually sampled from a nonlinear low dimensional manifold which is embedded in the high dimensional ambient space [4]. Yet NMF does not exploit the geometric structure of the data. In other word, it assumes that the data points are sampled from a Euclidean space. This greatly limits the application of NMF for the data lying on manifold. In order to consider the geometric structure in the data, we assume that if a data point can be reconstructed from its neighbors in the input space, then it can be reconstructed from its neighbors by the same reconstruction coefficients in the low dimensional subspace, i.e. *local linear embedding assumption* [4] [1].

For each data point $\mathbf{x}_i$, we use $\mathcal{N}_k(\mathbf{x}_i)$ to denote its $k$-nearest neighborhood. And we characterize the local geometric structure of its neighborhood by the linear coefficients that reconstruct $\mathbf{x}_i$ from its neighbors, i.e. $\mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i)$. The reconstruction coefficients are computed by the following objective function

$$\begin{aligned} \min \quad & ||\mathbf{x}_i - \sum_{\mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i)} M_{ij}\mathbf{x}_j||^2, \\ \text{s.t.} \quad & \sum_{\mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i)} M_{ij} = 1 \end{aligned} \tag{1}$$

And $M_{ij} = 0$ if $\mathbf{x}_j \notin \mathcal{N}_k(\mathbf{x}_i)$.

Then $\mathbf{v}_i, 1 \le i \le n$ in the low dimensional subspace can be reconstructed by minimizing

$$\begin{aligned} & \sum_i ||\mathbf{v}_i - \sum_{\mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i)} M_{ij}\mathbf{v}_j||^2 \\ = \; & \mathrm{tr}(\mathbf{V}(\mathbf{I}-\mathbf{M})(\mathbf{I}-\mathbf{M})\mathbf{V}^T) \\ = \; & \mathrm{tr}(\mathbf{V}\mathbf{L}\mathbf{V}^T) \end{aligned} \tag{2}$$

where $\mathbf{I} \in \mathbb{R}^{n\times n}$ is identity matrix and $\mathbf{L} = (\mathbf{I}-\mathbf{M})(\mathbf{I}-\mathbf{M})$. Eq.(2) is called *Neighborhood Preserving Regularization*. The better each point is reconstructed from its neighborhood in the low dimensional subspace, the smaller the neighborhood preserving regularizer will be.

Our assumption is that each point can be reconstructed by the data points in its neighborhood. To apply this idea for NMF, we constrain NMF with neighborhood preserving regularization in Eq.(2) as follows

$$\begin{aligned} J_{NPNMF} \; &= \; ||\mathbf{X}-\mathbf{U}\mathbf{V}||_F^2 + \mu\,\mathrm{tr}(\mathbf{V}\mathbf{L}\mathbf{V}^T), \\ & \text{s.t.} \quad \mathbf{U} \ge 0, \mathbf{V} \ge 0, \end{aligned} \tag{3}$$

where $\mu$ is a positive regularization parameter controlling the contribution of the additional constraint. We call Eq.(3) *Neighborhood Preserving Nonnegative Matrix Factorization* (NPNMF). Let $\mu = 0$, Eq.(3) degenerates to the original NMF. To make the objective in Eq.(3) lower bounded, we use $L_2$ normalization on rows of $\mathbf{V}$ in the optimization, and compensate the norms of $\mathbf{V}$ to $\mathbf{U}$.

In the following, we will give the solution to Eq.(3).

Since $\mathbf{U} \ge 0, \mathbf{V} \ge 0$, we introduce the Lagrangian multiplier $\gamma \in \mathbb{R}^{d\times m}$ and $\eta \in \mathbb{R}^{m\times n}$, thus, the Lagrangian function is

$$L(\mathbf{U},\mathbf{V}) = ||\mathbf{X}-\mathbf{U}\mathbf{V}||_F^2 + \mu\,\mathrm{tr}(\mathbf{V}\mathbf{L}\mathbf{V}^T) - \mathrm{tr}(\gamma\mathbf{U}^T) - \mathrm{tr}(\eta\mathbf{V}^T) \tag{4}$$

Setting $\frac{\partial L(\mathbf{U},\mathbf{V})}{\partial \mathbf{U}} = 0$ and $\frac{\partial L(\mathbf{U},\mathbf{V})}{\partial \mathbf{V}} = 0$, we obtain

$$\begin{aligned} \gamma &= -2\mathbf{X}\mathbf{V}^T + 2\mathbf{U}\mathbf{V}\mathbf{V}^T \\ \eta &= -2\mathbf{U}^T\mathbf{X} + 2\mathbf{U}^T\mathbf{U}\mathbf{V} + 2\mu\mathbf{V}\mathbf{L} \end{aligned} \tag{5}$$

Using the Karush-Kuhn-Tucker condition $\gamma_{ij}\mathbf{U}_{ij} = 0$ and $\eta_{ij}\mathbf{V}_{ij} = 0$, we get

$$\begin{aligned} (-\mathbf{X}\mathbf{V}^T + \mathbf{U}\mathbf{V}\mathbf{V}^T)_{ij}\mathbf{U}_{ij} &= 0 \\ (-\mathbf{U}^T\mathbf{X} + \mathbf{U}^T\mathbf{U}\mathbf{V} + \mu\mathbf{V}\mathbf{L})_{ij}\mathbf{V}_{ij} &= 0 \end{aligned} \tag{6}$$

Introduce

$$\mathbf{L} = \mathbf{L}^+ - \mathbf{L}^- \tag{7}$$

where $\mathbf{L}^+_{ij} = (|\mathbf{L}_{ij}| + \mathbf{L}_{ij})/2$ and $\mathbf{L}^-_{ij} = (|\mathbf{L}_{ij}| - \mathbf{L}_{ij})/2$.

Substitute Eq.(7) into Eq.(6), we obtain

$$\begin{aligned} (-\mathbf{X}\mathbf{V}^T + \mathbf{U}\mathbf{V}\mathbf{V}^T)_{ij}\mathbf{U}_{ij} &= 0 \\ (-\mathbf{U}^T\mathbf{X} + \mathbf{U}^T\mathbf{U}\mathbf{V} + \mu\mathbf{V}\mathbf{L}^+ - \mu\mathbf{V}\mathbf{L}^-)_{ij}\mathbf{V}_{ij} &= 0 \end{aligned} \tag{8}$$

Eq.(8) leads to the following updating formula

$$\begin{aligned} \mathbf{U}_{ij} &\leftarrow \mathbf{U}_{ij}\sqrt{\frac{(\mathbf{X}\mathbf{V}^T)_{ij}}{(\mathbf{U}\mathbf{V}\mathbf{V}^T)_{ij}}} \\ \mathbf{V}_{ij} &\leftarrow \mathbf{V}_{ij}\sqrt{\frac{(\mathbf{U}^T\mathbf{X} + \mu\mathbf{V}\mathbf{L}^-)_{ij}}{(\mathbf{U}^T\mathbf{U}\mathbf{V} + \mu\mathbf{V}\mathbf{L}^+)_{ij}}} \end{aligned} \tag{9}$$

Table 1 shows the experimental results of the methods on the ORL data set, where the value in each entry represents the average recognition accuracy of 20 independent trials, and the number in brackets is the corresponding projection dimensionality.

Table 1: Face Recognition accuracy on the ORL data set. The number in brackets is the corresponding projection dimensionality.

| Method | 2 Train | 3 Train | 4 Train |
|---|---|---|---|
| Baseline | 70.67 | 78.88 | 84.12 |
| PCA | 70.67(79) | 78.88(118) | 84.21(152) |
| LDA | 72.80(25) | 83.79(39) | 90.13(39) |
| NPE | 73.19(36) | 84.29(54) | 91.06(73) |
| NMF | 70.87(97) | 78.98(81) | 84.48(95) |
| LNMF | 71.73(178) | 81.09(168) | 86.31(195) |
| DNMF | 74.00(75) | 83.32(84) | 88.10(74) |
| NPNMF | **75.31(200)** | **84.73(94)** | **91.35(81)** |

We can see that our method outperforms other dimensionality reduction methods on the ORL data set. The superiority of our method may arise in the following two aspects: (1) *local linear embedding assumption* [4] [1], which preserves the local geometric structure of the data. (2) the *nonnegativity*, inheriting from NMF, which is suitable for nonnegative data, e.g. image data.

[1] Xiaofei He, Deng Cai, Shuicheng Yan, and HongJiang Zhang. Neighborhood preserving embedding. In *ICCV*, pages 1208–1213, 2005.

[2] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562, 2000.

[3] Stan Z. Li, XinWen Hou, HongJiang Zhang, and QianSheng Cheng. Learning spatially localized, parts-based representation. In *CVPR (1)*, pages 207–212, 2001.

[4] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, December 2000.

[5] Yuan Wang, Yunde Jia, Changbo Hu, and Matthew Turk. Fisher non-negative matrix factorization for learning local features. In *ACCV*, January 2004.