Subspace Clustering by Mixture of Gaussian Regression

Baohua Li[1],Ying Zhang[1], Zhouchen Lin[2,3] and Huchuan Lu[1]
[1]Dalian University of Technology. [2]Peking University. [3]Cooperative Medianet Innovation Center, Shanghai, China.

Subspace clustering is a problem of finding a multi-subspace representation that best fits sample points drawn from a high-dimensional space. Among the many approaches for subspace clustering, the spectral-clustering-based methods have shown excellent performance with two main steps. Firstly, an affinity matrix is built to capture the similarity between pairs of sample points. Secondly, graph cut is applied to a graph, whose vertices are the samples and whose weights are prescribed by the affinity matrix, for segmenting the sample points. Building a "good" affinity matrix is key to guarantee a good clustering result.

Given data matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N) \in \mathbb{R}^{M \times N}$ with N samples in $\mathbb{R}^M$, we denote $\boldsymbol{E} \in \mathbb{R}^{M \times N}$ and $\boldsymbol{Z} \in \mathbb{R}^{N \times N}$ as the noise matrix and the representation matrix, respectively, where the entry Z_{ij} of $\boldsymbol{Z}$ measures the similarity between points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

As described in [5], we consider subspace clustering as the following optimization problem:

$$\begin{aligned} &\min_{\boldsymbol{Z},\boldsymbol{E}} \mathcal{L}(\boldsymbol{E}) + \mathcal{R}(\boldsymbol{Z}) \\ &s.t.\ \boldsymbol{X} = \boldsymbol{X}\boldsymbol{Z} + \boldsymbol{E}, \end{aligned} \tag{1}$$

where $\mathcal{L}(\boldsymbol{E})$ is the loss function to describe noise and $\mathcal{R}(\boldsymbol{Z})$ is the regularization term to impose some properties on the representation matrix $\boldsymbol{Z}$.

Most previous spectral-clustering-based methods such as Sparse Subspace Clustering (SSC) [1], Low-Rank Representation (LRR) [2], Least Squares Regression (LSR) [3], Correlation Adaptive Subspace Segmentation (CASS) [4], Correntropy Induced L2 (CIL2) graph [5], rely on specific norms on $\boldsymbol{Z}$ and $\boldsymbol{E}$ to encourage the between-cluster sparsity and grouping effect of the representation matrix. Unfortunately, real noise in applications often exhibits very complex statistical distributions, rather than simply being Gaussian or sparse. So the noise cannot be easily described by a simple norm like the Frobenius norm, ℓ_1-norm, or $\ell_{2,1}$-norm.

To address this issue, we propose Mixture of Gaussian Regression (MoG Regression) for subspace clustering by modeling noise as a Mixture of Gaussians (MoG). The MoG Regression provides an effective way to model a much broader range of noise distributions. As a result, the obtained affinity matrix is better at characterizing the structure of data in real applications.

We assume that each column $\boldsymbol{e}_n\,(n = 1, \ldots, N)$ in $\boldsymbol{E}$ follows an MoG distribution, i.e.,

$$p(\boldsymbol{e}_n) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{e}_n \mid \boldsymbol{0}, \boldsymbol{\Sigma}_k), \tag{2}$$

where K is the number of Gaussian components and π_k denotes the mixing weight with $\pi_k \geq 0$ and $\sum_{k=1}^{K} \pi_k = 1$. $\mathcal{N}(\boldsymbol{e}_n \mid \boldsymbol{0}, \boldsymbol{\Sigma}_k)$ is the zero-mean multivariate Gaussian distribution, with $\boldsymbol{\Sigma}_k (k = 1, 2, \ldots, K)$ denoting the covariance matrix.

Similar to classical regression analysis, all columns in $\boldsymbol{E}$ are assumed to be independently and identically distributed. So we have

$$p(\boldsymbol{E}) = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{e}_n \mid \boldsymbol{0}, \boldsymbol{\Sigma}_k). \tag{3}$$

In a general MoG model, we wish to find $\boldsymbol{\pi} = (\pi_1, ..., \pi_K)^\top$ and $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_K)$ that maximize $p(\boldsymbol{E})$, which is equivalent to minimizing the negative log likelihood function defined as

$$-\ln p(\boldsymbol{E}) = -\sum_{n=1}^{N} \ln \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{e}_n \mid \boldsymbol{0}, \boldsymbol{\Sigma}_k) \right). \tag{4}$$

If we utilize $\mathcal{L}(\boldsymbol{E}) = -\ln p(\boldsymbol{E})$ to replace the Frobenius norm in the LSR model, then the proposed MoG Regression method can be formulated as follows:

$$\begin{aligned} &\min_{\boldsymbol{Z},\boldsymbol{E},\boldsymbol{\pi},\boldsymbol{\Sigma}} -\sum_{n=1}^{N} \ln \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{e}_n \mid \boldsymbol{0}, \boldsymbol{\Sigma}_k) \right) + \lambda \parallel \boldsymbol{Z} \parallel_F^2 \\ &s.t.\ \boldsymbol{X} = \boldsymbol{X}\boldsymbol{Z} + \boldsymbol{E}, \mathrm{diag}(\boldsymbol{Z}) = \boldsymbol{0}, \\ &\pi_k \geq 0, \boldsymbol{\Sigma}_k \in \mathbb{S}^+, k = 1, ..., K, \sum_{k=1}^{K} \pi_k = 1, \end{aligned} \tag{5}$$

where $\lambda > 0$ is the regularization parameter, $\mathbb{S}^+$ is the set of symmetrical positive definite (SPD) matrices and the constraint $\mathrm{diag}(\boldsymbol{Z}) = \boldsymbol{0}$ discourages using a sample to represent itself. Here we simply choose the Frobenius norm to regularize $\boldsymbol{Z}$.

After solving the MoG Regression problem (5) with EM algorithm to get the representation matrix $\boldsymbol{Z}$, we define the affinity matrix as

$$\boldsymbol{C} = \mid \boldsymbol{Z} \mid + \mid \boldsymbol{Z}^\top \mid,$$

where each entry C_{ij} in $\boldsymbol{C}$ measures the similarity between data points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. In the end, we employ Normalize Cut [6] on the affinity matrix $\boldsymbol{C}$ to produce the final clustering results.

Experimental results on multiple public databases show that the proposed method is effective and robust to noise in motion segmentation, handwritten digits clustering, and complex face clustering. Quantitative comparisons in Table 1 demonstrate that MoG Regression significantly outperforms state-of-the-art subspace clustering methods.

Table 1: The clustering accuracies (%) on the Hopkins 155 database, MNIST-Back-Rand database and AR databases.

Database		SSC	LRR	LSR	CASS	CIL2	Ours
Hopkins 155	2 motions	95.69	96.43	97.48	97.01	97.63	**98.76**
	3 motions	91.97	92.35	93.21	94.06	94.34	**95.03**
MNIST-Back-Rand		33.56	22.85	20.55	29.05	36.50	**51.98**
AR	5 subjects	83.05	84.41	87.69	78.46	85.38	**93.85**
	10 subjects	75.06	78.54	63.07	77.69	80.39	**88.85**

[1] Ehsan Elhamifar and Rene Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013.

[2] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.

[3] Can Yi Lu, Hai Min, Zhong Qiu Zhao, Lin Zhu, De Shuang Huang, and Shuicheng Yan. Robust and efficient subspace segmentation via least squares regression. In *Proceedings of European Conference on Computer Vision*, pages 347–360. 2012.

[4] Canyi Lu, Jiashi Feng, Zhouchen Lin, and Shuicheng Yan. Correlation adaptive subspace segmentation by trace lasso. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1345–1352, 2013.

[5] Canyi Lu, Jinhui Tang, Min Lin, Liang Lin, Shuicheng Yan, and Zhouchen Lin. Correntropy induced l2 graph for robust subspace clustering. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1801–1808, 2013.

[6] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.