Low-Rank Correlation Analysis

Jiacan Zheng

College of Computer Science and Software Engineering
Shenzhen University

2021/11/8

Context and Motivation

Context:

To avoid the curse of dimensionality problem, a commonly used technique is linear dimensionality reduction. However, learning a discriminative subspace without label is still a challenging problem, especially when the high-dimensional data is grossly corrupted.

Motivation:

Recently, some innovative works further demonstrate the great potential of the multi-subspace structure prior in representation learning. Motivated by these encouraging researches, our model **LRCA** focus on enforcing the multi-subspace structure into the linear embedding space to improve the discriminative ability.

LRCA: self-expressive data matrix DZ

Given the data matrix $\mathbf{X}$, we first optimze the self-expressive matrix $\mathbf{Z}$ by LRR:

$$\min_{\mathbf{Z}} \ \|\mathbf{E}\|_{2,1} + \lambda \|\mathbf{Z}\|_*, \quad \text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}. \tag{1}$$

Suppose we have an over-complete dictionary $\mathbf{D} \in \mathbb{R}^{k \times n}$, we then construct the so-called self-expressive data matrix $\mathbf{DZ}$.

As the coefficient matrix $\mathbf{Z}$ captures the multi-subspace structure of the original data, self-expressive data matrix $\mathbf{DZ}$ then have the potential to preserve such a structure into the embedding space.

LRCA: maximizing the correlation

We maximize the correlation criterion between the linear embedding $\mathbf{P}^T\mathbf{X}$ and the self-expressive data matrix $\mathbf{DZ}$ to find a latent correlated subspace:

$$\mathrm{Corr}(\mathbf{D}, \mathbf{P}) = \frac{\langle \mathbf{DZ}, \mathbf{P}^T\mathbf{X} \rangle}{\sqrt{\langle \mathbf{DZ}, \mathbf{DZ} \rangle \langle \mathbf{P}^T\mathbf{X}, \mathbf{P}^T\mathbf{X} \rangle}}. \tag{2}$$

In order to be equipped with some regularizations, we modified the correlation criterion as:

$$\max_{\mathbf{D},\mathbf{P}} \frac{\mathcal{J}(\mathbf{D}, \mathbf{P})}{\sqrt{\mathrm{tr}\left(\mathbf{D}\ \mathbf{D}^T\right)}}, \tag{3}$$

where $= \mathbf{ZZ}^T$, and $\mathcal{J}(\mathbf{D}, \mathbf{P})$ is defined as:

$$\mathcal{J}(\mathbf{D}, \mathbf{P}) = 2\langle \mathbf{DZ}, \mathbf{P}^T\mathbf{X} \rangle - \mathrm{tr}(\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}) - \alpha\mathcal{R}(\mathbf{P}). \tag{4}$$

LRCA: Locality-preserving Dictionary D

We want to optimize a locality-preserving dictionary $\mathbf{D}$:

$$\min_{\mathbf{D}} \sum_j \sum_i \|\mathbf{d}_i - \mathbf{d}_j\|_2^2 \mathbf{W}_{ij} \Leftrightarrow \min_{\mathbf{D}} \ \mathrm{tr}(\mathbf{D}\mathbf{L}\mathbf{D}^T), \tag{5}$$

where $\mathbf{W}$ is an affinity matrix (constructed by $\mathbf{Z}$) and $\mathbf{L}$ is the corresponding Laplacian matrix.

To facilitate optimization, by fixing the denominator, we can optimize the constraint correlation criterion as follows:

$$\begin{aligned} &\max_{\mathbf{D},\mathbf{P}} \ \mathcal{J}(\mathbf{D},\mathbf{P}) \\ &\text{s.t. } \mathbf{D}\mathbf{L}\mathbf{D}^T = \mathbf{I}_{\mathbf{k}}. \end{aligned} \tag{6}$$

LRCA: A joint Optimization Formulation

we design a joint formulation, called Low-Rank Correlation Analysis (LRCA), as follows:

$$\begin{aligned} \min_{\mathbf{E},\mathbf{Z},\mathbf{D},\mathbf{P}} \ & \|\mathbf{E}\|_{2,1} + \lambda\|\mathbf{Z}\|_* - \gamma\mathcal{J}(\mathbf{D},\mathbf{P}) \\ \text{s.t. } & \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E},\ \mathbf{D}\mathbf{L}\mathbf{D}^T = \mathbf{I}_\mathbf{k}. \end{aligned} \tag{7}$$

Optimization (ADMM + eigen-decomposition):

Algorithm 2: LOW RANK CORRELATION ANALYSIS.

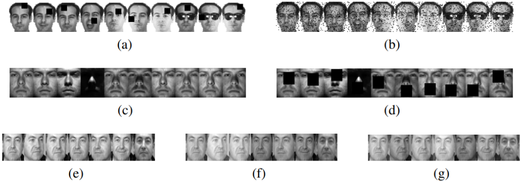
Input: Data matrix $\mathbf{X}$, parameter λ, γ, α, iteration times T

Output: Dictionary $\mathbf{D}$, Projection matrix $\mathbf{P}$

1 Initialize coefficient matrix $\mathbf{Z}$ by LRR
2 **for** $i = 1 : T$ **do**
3   Update $\mathbf{D}$ by (20).
4   Update $\mathbf{P}$ by (18).
5   Compute $\mathbf{M} = \mathbf{D}^T\mathbf{P}^T\mathbf{X}$.
6   Update $(\mathbf{Z}, \mathbf{E})$ by Algorithm 1.

Experiments

Datasets:



Classifcation Performance:

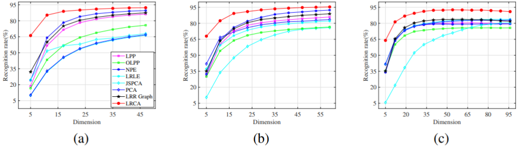


Fig. 2. Recognition on the (a) Extended Yale B (T = 30) (b) AR (T = 5) and (c) AR (10% salt-and-pepper noise, T = 7).

SHENZHEN UNIVERSITY

Experiments

Visualization of the learned subspace:

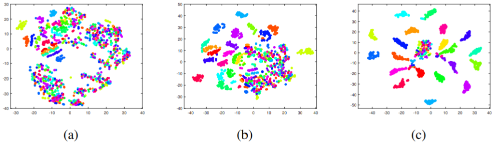


Fig. 4. t-SNE representation of the first 40 dimensions obtained by (a) PCA, (b) LPP and (c) LRCA on Extended Yale B dataset (T = 30).

Convergence of the algorithm:

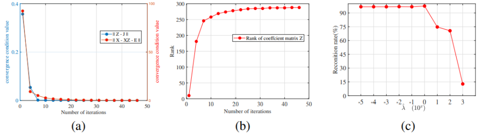


Fig. 5. The variation of (a) two convergence conditions of ADMM Algorithm 1, (b) the corresponding coefficient matrix **Z**'s rank on Extend Yale B dataset. (c) Recognition rates versus the value of parameter $\hat{\lambda}$ on AR dataset.

Q & A

Thanks for Listening!

Email: JiacanZheng@qq.com