Positive Collaborative Representation for Subspace Clustering

June 14, 2017

Abstract

1 Introduction

2 Motivation

- Positive collaborative representation could achieve sparse representation since similar points are sparse while dissimilar points are dense.

- Positive supports are positive to self-reppresentation while negative supports are negative to self-representation.

- Better performance. Faster?

3 LSR Model

The least squares regression (LSR) model [1] is proposed by Lu et al. can be formulated as follows:

$$\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_F^2 \text{ s.t. } \mathrm{diag}(\boldsymbol{A}) = \boldsymbol{0}. \quad (1)$$

Here we denote by $\mathrm{diag}(\boldsymbol{A})$ both a diagonal matrix whose diagonal elements are the diagonal entries of $\boldsymbol{A}$ and the vector consisted of the diagonal elements. According to [1], the above problem has the optimal solution as

$$\hat{\boldsymbol{A}} = -\boldsymbol{Z}(\mathrm{diag}(\boldsymbol{Z})) \text{ s.t. } \mathrm{diag}(\hat{\boldsymbol{A}}) = \boldsymbol{0}, \quad (2)$$

where $\boldsymbol{Z} = (\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}$.

The constraint of $\mathrm{diag}(\boldsymbol{A}) = \boldsymbol{0}$ in (1) could be removed and the LSR model achieves similar performance.

4 Collaborative Representation based Clustering with Constraint $\mathrm{diag}(A) = 0$

The LSR model can be reformulated as a collaborative representation model [2] for subspace clustering with an additional constraint of $\mathrm{diag}(\boldsymbol{A}) = \boldsymbol{0}$. The constraint of $\mathrm{diag}(\boldsymbol{A}) = \boldsymbol{0}$ is used to avoid the samples to represent themselves.

By introducing auxiliary variables into the optimization program, we can set $\boldsymbol{C} = \boldsymbol{A}$. The LSR model (1) can be transformed into

$$\begin{aligned} &\min_{\boldsymbol{A},\boldsymbol{C}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^F + \lambda\|\boldsymbol{C}\|_F^2 \\ &\text{s.t. } \boldsymbol{C} = \boldsymbol{A} - \mathrm{diag}(\boldsymbol{A}), \end{aligned} \quad (3)$$

whose solution for $\boldsymbol{A}$ coincides with the solution of Eq. (1). By introducing a Lagrangian multipliers $\boldsymbol{\Delta}$ and a penalty parameter ρ, the Lagrangian function of the Eq. (29) can be written as

$$\begin{aligned} &\mathcal{L}(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\Delta}, \rho) = \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^F + \lambda\|\boldsymbol{C}\|_F^2 \\ &+ \langle \boldsymbol{\Delta}, \boldsymbol{C} - (\boldsymbol{A} - \mathrm{diag}(\boldsymbol{A}))\rangle + \frac{\rho}{2}\|\boldsymbol{C} - (\boldsymbol{A} - \mathrm{diag}(\boldsymbol{A}))\|_F^2 \end{aligned} \quad (4)$$

Denote by $(\boldsymbol{C}_k, \boldsymbol{A}_k)$ the optimization variables at iteration k, by $\boldsymbol{\Delta}_k$ the Lagrangian multipliers at iteration k, and by ρ the penalty parameter at iteration k. Taking detivatives of $\mathcal{L}$ with respect to the variables and setting the derivatives to be zeros, we can alternatively update the variables as follows:

(1) Obtain $\boldsymbol{A}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{A}$, while fixing $(\boldsymbol{C}_k, \boldsymbol{\Delta}_k)$. This is equivalent to solve the fol-

lowing problem:

$$\begin{aligned} &\boldsymbol{A}_{k+1} = \boldsymbol{J} - \text{diag}(\boldsymbol{J}), \\ &\boldsymbol{J} = (\boldsymbol{X}^\top \boldsymbol{X} + \frac{\rho}{2}\boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + \frac{\rho}{2}\boldsymbol{C}_k + \frac{1}{2}\boldsymbol{\Delta}_k) \end{aligned} \tag{5}$$

(2) Obtain $\boldsymbol{C}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{C}$, while fixing $(\boldsymbol{A}_{k+1}, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$\boldsymbol{C}_{k+1} = \arg\min_{\boldsymbol{C}} \frac{\rho}{2}\|\boldsymbol{C} - (\boldsymbol{A}_{k+1} - \rho^{-1}\boldsymbol{\Delta}_k)\|_F^2 + \lambda\|\boldsymbol{C}\|_F^2 \tag{6}$$

This is a least squares regression problem which has a closed-form solution as

$$\boldsymbol{C}_{k+1} = (\rho + 2\lambda)^{-1}(\rho\boldsymbol{A}_{k+1} - \boldsymbol{\Delta}_k). \tag{7}$$

(3) Obtain the Lagrangian multipliers $\boldsymbol{\Delta}_{k+1}$ while fixing $(\boldsymbol{C}_{k+1}, \boldsymbol{A}_{k+1})$:

$$\boldsymbol{\Delta}_{k+1} = \boldsymbol{\Delta}_k + \tau\rho(\boldsymbol{C}_{k+1} - \boldsymbol{A}_{k+1}), \tag{8}$$

where $\tau \in (0, \frac{\sqrt{5}+1}{2})$ is the dual step size and is usually set as $\tau = 1$.

Convergency analysis?

5 Non-Negatie Collaborative Representation

This model enforces non-negative representation and hence produce sparse solutions, in the sense that it results only a few non-negative coefficients.

The performance of this method is much better than the original least squares regression (LSR) based subspace clustering method proposed by Lu et al. [1].

6 Large Scale Subset Selection Via Woodbury Identity

The Woodbury Identity is

$$(\boldsymbol{A}+\boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}. \tag{9}$$

We can also restrict that $\text{diag}(\boldsymbol{A}) = \boldsymbol{0}$ to avoid the samples to be self-represented. However, I want to mention that the proposed model solved by ADMM algorithm with three variables and does not have convergence results.

Then the model above can be

$$\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_{p,1}. \tag{10}$$

By introducing an auxiliary variable $\boldsymbol{C}$ into the optimization program, we can get

$$\min_{\boldsymbol{A},\boldsymbol{C}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{C}\|_{p,1} \text{ s.t. } \boldsymbol{C} = \boldsymbol{A}, \tag{11}$$

whose solution for $\boldsymbol{A}$ coincides with the solution of Eq. (29). By introducing two Lagrangian multipliers $\boldsymbol{\Delta}$, the Lagrangian function of the Eq. (29) can be written as

$$\begin{aligned} &\mathcal{L}(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\Delta}, \rho) = \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{C}\|_{p,1} \\ &+ \langle \boldsymbol{\Delta}, \boldsymbol{C} - \boldsymbol{A} \rangle + \frac{\rho}{2}\|\boldsymbol{C} - \boldsymbol{A}\|_F^2 \end{aligned} \tag{12}$$

Denote by $(\boldsymbol{C}_k, \boldsymbol{A}_k)$ the optimization variables at iteration k, by $\boldsymbol{\Delta}_k$ the Lagrangian multipliers at iteration k, and by ρ the penalty parameter at iteration k. Taking detivatives of $\mathcal{L}$ with respect to the variables and setting the derivatives to be zeros, we can alternatively update the variables as follows:

(1) Obtain $\boldsymbol{A}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{A}$, while fixing $(\boldsymbol{C}_k, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \frac{\rho}{2}\|\boldsymbol{A} - (\boldsymbol{C}_k + \rho^{-1}\boldsymbol{\Delta}_k)\|_F^2, \tag{13}$$

which is equivalently to solve the following problem

$$\boldsymbol{A} = (\boldsymbol{X}^\top \boldsymbol{X} + \frac{\rho}{2}\boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + \frac{\rho}{2}\boldsymbol{C}_k + \frac{1}{2}\boldsymbol{\Delta}_k) \tag{14}$$

Since the matrices $\boldsymbol{X}^\top \boldsymbol{X}$ is of $N \times N$ dimension. It is computational expensive when N is very large. By employing the Woodburry Identity mentioned above, we can have

$$(\frac{\rho}{2}\boldsymbol{I} + \boldsymbol{X}^\top \boldsymbol{X})^{-1} = \frac{2}{\rho}\boldsymbol{I} - (\frac{2}{\rho})^2\boldsymbol{X}^\top(\boldsymbol{I} + \frac{2}{\rho}\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{X}. \tag{15}$$

and transform this problem as

$$\begin{aligned} \boldsymbol{A} = &(\frac{2}{\rho}\boldsymbol{I} - (\frac{2}{\rho})^2\boldsymbol{X}^\top(\boldsymbol{I} + \frac{2}{\rho}\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{X}) \\ &* (\boldsymbol{X}^\top \boldsymbol{X} + \frac{\rho}{2}\boldsymbol{C}_k + \frac{1}{2}\boldsymbol{\Delta}_k) \end{aligned} \tag{16}$$

which will save a lot of computational costs.

(2) Obtain $\boldsymbol{C}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{C}$, while fixing $(\boldsymbol{A}_{k+1}, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$\min_{\boldsymbol{C}} \frac{1}{2}\|(\boldsymbol{A}_{k+1} - \rho^{-1}\boldsymbol{\Delta}k) - \boldsymbol{C}\|_F^2 + \frac{\lambda}{\rho}\|\boldsymbol{C}\|_{p,1}. \quad (17)$$

Since the $\ell_{p,1}$ norm is separable with respect to each row, we can write the above problem as

$$\min_{\boldsymbol{C}} \sum_{i=1}^{M} \frac{1}{2}\|(\boldsymbol{A}_{k+1})_{i*} - \rho^{-1}(\boldsymbol{\Delta}_k)_{i*} - \boldsymbol{C}_{i*}\|_2^2 + \frac{\lambda}{\rho}\|\boldsymbol{C}_{i*}\|_p, \quad (18)$$

where $\boldsymbol{F}_{i*}$ is the ith row of the matrix $\boldsymbol{F}$. Since this step is separable w.r.t. each row, we can employ parallel processing resources and reduce its computational time.

(3) Obtain the Lagrangian multipliers $(\boldsymbol{\Delta}_{k+1})$ while fixing $(\boldsymbol{C}_{k+1}, \boldsymbol{A}_{k+1})$:

$$\boldsymbol{\Delta}_{k+1} = \boldsymbol{\Delta}_k + \rho(\boldsymbol{C}_{k+1} - \boldsymbol{A}_{k+1}). \quad (19)$$

(5) Update the penalty parameter ρ as $\rho = \mu\rho$, where $\mu > 1$.

7 Robust Large Scale Subset Selection via Dissimilarity based Outlier Detection

We can also introduce a dissimiarlity based matrix $\boldsymbol{D}$ to replace the ℓ_p or $\ell_{2,1}$ norms to ensure robustness. This can also remore the additional term $\boldsymbol{Z}$ on modeling the outliers with the restriction of ℓ_1 norm. The matrix $\boldsymbol{D}$ should better be diagonal matrix. How to design the matrix $\boldsymbol{D}$ is another problem need to be solved.

Then the proposed model can be formulated as

$$\min_{\boldsymbol{A}} \|(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A})\boldsymbol{D}\|_F^2 + \lambda\|\boldsymbol{A}\|_{p,1}. \quad (20)$$

By introducing an auxiliary variable $\boldsymbol{C}$ into the optimization program, we can get

$$\min_{\boldsymbol{A},\boldsymbol{C}} \|(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A})\boldsymbol{D}\|_F^2 + \lambda\|\boldsymbol{C}\|_{p,1} \text{ s.t. } \boldsymbol{C} = \boldsymbol{A}. \quad (21)$$

By introducing a Lagrangian multiplier $\boldsymbol{\Delta}$, the Lagrangian function of the Eq. (29) can be written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\Delta}, \rho) &= \|(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A})\boldsymbol{D}\|_F^2 + \lambda\|\boldsymbol{C}\|_{p,1} \\ &+ \langle \boldsymbol{\Delta}, \boldsymbol{C} - \boldsymbol{A}\rangle + \frac{\rho}{2}\|\boldsymbol{C} - \boldsymbol{A}\|_F^2\end{aligned} \quad (22)$$

Denote by $(\boldsymbol{A}_k, \boldsymbol{C}_k)$ the optimization variables at iteration k, by $\boldsymbol{\Delta}_k$ the Lagrangian multiplier at iteration k, and by ρ the penalty parameter at iteration k. Taking detivatives of $\mathcal{L}$ with respect to the variables and setting the derivatives to be zeros, we can alternatively update the variables as follows:

(1) Obtain $\boldsymbol{A}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{A}$, while fixing $(\boldsymbol{C}_k, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$\min_{\boldsymbol{A}} \|(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A})\boldsymbol{D}\|_F^2 + \frac{\rho}{2}\|\boldsymbol{A} - (\boldsymbol{C}_k - \rho^{-1}\boldsymbol{\Delta}_k)\|_F^2, \quad (23)$$

which is equalivalently to solve the following problem

$$\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{A}\boldsymbol{D}\boldsymbol{D}^\top + \frac{\rho}{2}\boldsymbol{A} = \boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{D}\boldsymbol{D}^\top + \frac{\rho}{2}(\boldsymbol{C}_k - \rho^{-1}\boldsymbol{\Delta}_k) \quad (24)$$

Since the matrices $\boldsymbol{X}^\top\boldsymbol{X}$ and $\boldsymbol{D}^\top\boldsymbol{D}$ are positive semidefinite and positive definite, respectively. The above equation is a standard Sylvester equation which has a unique solution.

(2) Obtain $\boldsymbol{C}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{C}$, while fixing $(\boldsymbol{A}_{k+1}, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$\min_{\boldsymbol{C}} \frac{1}{2}\|(\boldsymbol{A}_{k+1} + \rho^{-1}\boldsymbol{\Delta}k) - \boldsymbol{C}\|_F^2 + \frac{\lambda}{\rho}\|\boldsymbol{C}\|_{p,1}. \quad (25)$$

Since the $\ell_{p,1}$ norm is separable with respect to each row, we can write the above problem as

$$\min_{\boldsymbol{C}} \sum_{i=1}^{M} \frac{1}{2}\|(\boldsymbol{A}_{k+1})_{i*} + \rho^{-1}(\boldsymbol{\Delta}_k)_{i*} - \boldsymbol{C}_{i*}\|_2^2 + \frac{\lambda}{\rho}\|\boldsymbol{C}_{i*}\|_p, \quad (26)$$

where $\boldsymbol{F}_{i*}$ is the ith row of the matrix $\boldsymbol{F}$. Since this step is separable w.r.t. each row, we can employ parallel processing resources and reduce its computational time.

(3) Obtain the Lagrangian multipliers ($\boldsymbol{\Delta}_{k+1}$) while fixing ($\boldsymbol{C}_{k+1}, \boldsymbol{A}_{k+1}$):

$$\boldsymbol{\Delta}_{k+1} = \boldsymbol{\Delta}_k + \rho(\boldsymbol{C}_{k+1} - \boldsymbol{A}_{k+1}). \tag{27}$$

(5) Update the penalty parameter ρ as $\rho = \mu\rho$, where $\mu > 1$.

8 Large Scale Subset Selection Via Row-Column Separation

We can also restrict that $\mathrm{diag}(\boldsymbol{A}) = \mathbf{0}$ to avoid the samples to be self-represented. However, I want to mention that the proposed model solved by ADMM algorithm with three variables and does not have convergence results.

Then the model above can be

$$\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_{p,1} \quad \text{s.t.} \quad \mathrm{diag}(\boldsymbol{A}) = \mathbf{0}. \tag{28}$$

By introducing an auxiliary variable $\boldsymbol{C}$ into the optimization program, we can get

$$\begin{aligned} &\min_{\boldsymbol{A},\boldsymbol{C}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}\|_F^2 + \lambda\|\boldsymbol{A}\|_{p,1} \\ &\text{s.t. } \boldsymbol{C} = \boldsymbol{A} - \mathrm{diag}(\boldsymbol{A}), \end{aligned} \tag{29}$$

whose solution for $\boldsymbol{A}$ coincides with the solution of Eq. (29). By introducing two Lagrangian multipliers $\boldsymbol{\Delta}$, the Lagrangian function of the Eq. (29) can be written as

$$\begin{aligned} &\mathcal{L}(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\Delta}, \rho) = \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}\|_F^2 + \lambda\|\boldsymbol{A}\|_{p,1} \\ &+ \langle \boldsymbol{\Delta}, \boldsymbol{C} - (\boldsymbol{A} - \mathrm{diag}(\boldsymbol{A})) \rangle + \frac{\rho}{2}\|\boldsymbol{C} - (\boldsymbol{A} - \mathrm{diag}(\boldsymbol{A})\|_F^2 \end{aligned} \tag{30}$$

Denote by ($\boldsymbol{C}_k, \boldsymbol{A}_k$) the optimization variables at iteration k, by $\boldsymbol{\Delta}_k$ the Lagrangian multipliers at iteration k, and by ρ the penalty parameter at iteration k. Taking detivatives of $\mathcal{L}$ with respect to the variables and setting the derivatives to be zeros, we can alternatively update the variables as follows:

(1) Obtain $\boldsymbol{A}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{A}$, while fixing ($\boldsymbol{C}_k, \boldsymbol{\Delta}_k$). This is equivalent to solve the following problem:

$$\begin{aligned} &\boldsymbol{A}_{k+1} = \boldsymbol{J} - \mathrm{diag}(\boldsymbol{J}), \\ &\boldsymbol{J} = \arg\min_{\boldsymbol{J}} \frac{1}{2}\|\boldsymbol{C}_k + \rho^{-1}\boldsymbol{\Delta}_k - \boldsymbol{J}\|_F^2 + \frac{\lambda}{\rho}\|\boldsymbol{J}\|_{p,1}. \end{aligned} \tag{31}$$

(2) Obtain $\boldsymbol{C}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{C}$, while fixing ($\boldsymbol{A}_{k+1}, \boldsymbol{\Delta}_k$). This is equivalent to solve the following problem:

$$\min_{\boldsymbol{C}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}\|_F^2 + \frac{\rho}{2}\|\boldsymbol{C} - \boldsymbol{A}_{k+1} + \frac{1}{\rho}\boldsymbol{\Delta}_k\|_F^2 \tag{32}$$

This is a least squares regression problem which has a closed-form solution as

$$\boldsymbol{C}_{k+1} = (\boldsymbol{X}^\top\boldsymbol{X} + \frac{\rho}{2}\boldsymbol{I})^{-1}(\boldsymbol{X}^\top\boldsymbol{X} + \frac{\rho}{2}\boldsymbol{A}_{k+1} - \frac{1}{2}\boldsymbol{\Delta}_k). \tag{33}$$

(3) Obtain the Lagrangian multipliers ($\boldsymbol{\Delta}_{k+1}$) while fixing ($\boldsymbol{C}_{k+1}, \boldsymbol{A}_{k+1}$):

$$\boldsymbol{\Delta}_{k+1} = \boldsymbol{\Delta}_k + \rho(\boldsymbol{C}_{k+1} - \boldsymbol{A}_{k+1}). \tag{34}$$

(5) Update the penalty parameter ρ as $\rho = \mu\rho$, where $\mu > 1$.

References

[1] 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. *ECCV*, pages 347–360, 2012. 1

[2] Lei Zhang, Meng Yang, and Xiangchu Feng. Sparse representation or collaborative representation: Which helps face recognition? *ICCV*, pages 471–478, 2011. 1