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:

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

Here we denote by $\text{diag}(\boldsymbol{A})$ 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

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

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

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

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

The LSR model can be reformulated as a collaborative representation model [2] for subspace clustering with an additional constraint of $\text{diag}(\boldsymbol{A}) = \boldsymbol{0}$. The constraint of $\text{diag}(\boldsymbol{A}) = \boldsymbol{0}$ 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

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

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. (31) can be written as

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

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:

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

(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:

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

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

$$\begin{aligned}\boldsymbol{C}_{k+1}=\boldsymbol{H}-\mathrm{diag}(\boldsymbol{H}),\\(\boldsymbol{X}^\top\boldsymbol{X}+\boldsymbol{I})\boldsymbol{H}=\boldsymbol{X}^\top\boldsymbol{P}_k+\boldsymbol{Q}_k,\end{aligned}\tag{7}$$

where $\boldsymbol{P}_k=\boldsymbol{X}-\boldsymbol{E}_k-\rho^{-1}\boldsymbol{\Delta}_k$ and $\boldsymbol{Q}_k=\boldsymbol{A}_{k+1}-\rho^{-1}\boldsymbol{\delta}_k$.

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

$$\min_{\boldsymbol{E}}\frac{1}{2}\|(\boldsymbol{X}-\boldsymbol{X}\boldsymbol{C}_{k+1}-\rho^{-1}\boldsymbol{\Delta}_k)-\boldsymbol{E}\|_F^2+\rho^{-1}\|\boldsymbol{E}\|_1.\tag{8}$$

The solution of $\boldsymbol{E}$ can be computed in closed-form as

$$\boldsymbol{E}_{k+1}=\mathcal{S}_{\rho^{-1}}(\boldsymbol{X}-\boldsymbol{X}\boldsymbol{C}_{k+1}-\rho^{-1}\boldsymbol{\Delta}_k),\tag{9}$$

where $\mathcal{S}_\tau(x)=\mathrm{sign}(x)*\max(|x|-\tau,0)$ is the soft-thresholding operator.

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

$$\begin{aligned}\boldsymbol{\Delta}_{k+1}&=\boldsymbol{\Delta}_k+\tau\rho(\boldsymbol{E}_{k+1}-\boldsymbol{X}+\boldsymbol{X}\boldsymbol{C}_{k+1}),\\\boldsymbol{\delta}_{k+1}&=\boldsymbol{\delta}_k+\tau\rho(\boldsymbol{C}_{k+1}-\boldsymbol{A}_{k+1}),\end{aligned}\tag{10}$$

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

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

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.

6 Large Scale Subset Selection Via Woodbury Identity

The Woodbury Identity is

$$(\boldsymbol{A}+\boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}.\tag{11}$$

We can also restrict that $\mathrm{diag}(\boldsymbol{A})=\boldsymbol{0}$ 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

$$\min_{\boldsymbol{A}}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{A}\|_F^2+\lambda\|\boldsymbol{A}\|_{p,1}.\tag{12}$$

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

$$\min_{\boldsymbol{A},\boldsymbol{C}}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{A}\|_F^2+\lambda\|\boldsymbol{C}\|_{p,1}\ \text{s.t.}\ \boldsymbol{C}=\boldsymbol{A},\tag{13}$$

whose solution for $\boldsymbol{A}$ coincides with the solution of Eq. (31). By introducing two Lagrangian multipliers $\boldsymbol{\Delta}$, the Lagrangian function of the Eq. (31) can be written as

$$\begin{aligned}&\mathcal{L}(\boldsymbol{A},\boldsymbol{C},\boldsymbol{\Delta},\rho)=\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{A}\|_F^2+\lambda\|\boldsymbol{C}\|_{p,1}\\&+\langle\boldsymbol{\Delta},\boldsymbol{C}-\boldsymbol{A}\rangle+\frac{\rho}{2}\|\boldsymbol{C}-\boldsymbol{A}\|_F^2\end{aligned}\tag{14}$$

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:

$$\min_{\boldsymbol{A}}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{A}\|_F^2+\frac{\rho}{2}\|\boldsymbol{A}-(\boldsymbol{C}_k+\rho^{-1}\boldsymbol{\Delta}_k)\|_F^2,\tag{15}$$

which is equalivalently to solve the following problem

$$\boldsymbol{A} = (\boldsymbol{X}^\top \boldsymbol{X} + \frac{\rho}{2}\boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + \frac{\rho}{2}\boldsymbol{C}_k + \frac{1}{2}\boldsymbol{\Delta}_k) \quad (16)$$

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

$$(\frac{\rho}{2}\boldsymbol{I} + \boldsymbol{X}^\top \boldsymbol{X})^{-1} = \frac{2}{\rho}\boldsymbol{I} - (\frac{2}{\rho})^2 \boldsymbol{X}^\top (\boldsymbol{I} + \frac{2}{\rho}\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{X}. \quad (17)$$

and transform this problem as

$$\begin{aligned} \boldsymbol{A} =& (\frac{2}{\rho}\boldsymbol{I} - (\frac{2}{\rho})^2 \boldsymbol{X}^\top (\boldsymbol{I} + \frac{2}{\rho}\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{X}) \\ & * (\boldsymbol{X}^\top \boldsymbol{X} + \frac{\rho}{2}\boldsymbol{C}_k + \frac{1}{2}\boldsymbol{\Delta}_k) \end{aligned} \quad (18)$$

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:

$$\min_{\boldsymbol{C}} \frac{1}{2}\|(\boldsymbol{A}_{k+1} - \rho^{-1}\boldsymbol{\Delta}k) - \boldsymbol{C}\|_F^2 + \frac{\lambda}{\rho}\|\boldsymbol{C}\|_{p,1}. \quad (19)$$

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

$$\min_{\boldsymbol{C}} \sum_{i=1}^{M} \frac{1}{2}\|(\boldsymbol{A}_{k+1})_{i*} - \rho^{-1}(\boldsymbol{\Delta}_k)_{i*} - \boldsymbol{C}_{i*}\|_2^2 + \frac{\lambda}{\rho}\|\boldsymbol{C}_{i*}\|_p, \quad (20)$$

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 $(\boldsymbol{C}_{k+1}, \boldsymbol{A}_{k+1})$:

$$\boldsymbol{\Delta}_{k+1} = \boldsymbol{\Delta}_k + \rho(\boldsymbol{C}_{k+1} - \boldsymbol{A}_{k+1}). \quad (21)$$

(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

$$\min_{\boldsymbol{A}} \|(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A})\boldsymbol{D}\|_F^2 + \lambda\|\boldsymbol{A}\|_{p,1}. \quad (22)$$

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

$$\min_{\boldsymbol{A},\boldsymbol{C}} \|(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A})\boldsymbol{D}\|_F^2 + \lambda\|\boldsymbol{C}\|_{p,1} \text{ s.t. } \boldsymbol{C} = \boldsymbol{A}. \quad (23)$$

By introducing a Lagrangian multiplier $\boldsymbol{\Delta}$, the Lagrangian function of the Eq. (31) can be written as

$$\begin{aligned} &\mathcal{L}(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\Delta}, \rho) = \|(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A})\boldsymbol{D}\|_F^2 + \lambda\|\boldsymbol{C}\|_{p,1} \\ &+ \langle \boldsymbol{\Delta}, \boldsymbol{C} - \boldsymbol{A}\rangle + \frac{\rho}{2}\|\boldsymbol{C} - \boldsymbol{A}\|_F^2 \end{aligned} \quad (24)$$

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:

$$\min_{\boldsymbol{A}} \|(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A})\boldsymbol{D}\|_F^2 + \frac{\rho}{2}\|\boldsymbol{A} - (\boldsymbol{C}_k - \rho^{-1}\boldsymbol{\Delta}_k)\|_F^2, \quad (25)$$

which is equalivalently to solve the following problem

$$\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{A}\boldsymbol{D}\boldsymbol{D}^\top + \frac{\rho}{2}\boldsymbol{A} = \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{D}\boldsymbol{D}^\top + \frac{\rho}{2}(\boldsymbol{C}_k - \rho^{-1}\boldsymbol{\Delta}_k) \quad (26)$$

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 C_{k+1} by minimizing $\mathcal{L}$ with respect to C, while fixing (A_{k+1}, Δ_k). This is equivalent to solve the following problem:

$$\min_{C} \frac{1}{2}\|(A_{k+1} + \rho^{-1}\Delta k) - C\|_F^2 + \frac{\lambda}{\rho}\|C\|_{p,1}. \quad (27)$$

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

$$\min_{C} \sum_{i=1}^{M} \frac{1}{2}\|(A_{k+1})_{i*} + \rho^{-1}(\Delta_k)_{i*} - C_{i*}\|_2^2 + \frac{\lambda}{\rho}\|C_{i*}\|_p, \quad (28)$$

where F_{i*} is the ith row of the matrix 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 (Δ_{k+1}) while fixing (C_{k+1}, A_{k+1}):

$$\Delta_{k+1} = \Delta_k + \rho(C_{k+1} - A_{k+1}). \quad (29)$$

(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 $\text{diag}(A) = \mathbf{0}$ 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

$$\min_{A} \|X - XA\|_F^2 + \lambda\|A\|_{p,1} \quad \text{s.t.} \quad \text{diag}(A) = \mathbf{0}. \quad (30)$$

By introducing an auxiliary variable C into the optimization program, we can get

$$\begin{aligned} &\min_{A,C} \|X - XC\|_F^2 + \lambda\|A\|_{p,1} \\ &\text{s.t. } C = A - \text{diag}(A), \end{aligned} \quad (31)$$

whose solution for A coincides with the solution of Eq. (31). By introducing two Lagrangian multipliers Δ, the Lagrangian function of the Eq. (31) can be written as

$$\begin{aligned} &\mathcal{L}(A, C, \Delta, \rho) = \|X - XC\|_F^2 + \lambda\|A\|_{p,1} \\ &+ \langle \Delta, C - (A - \text{diag}(A)) \rangle + \frac{\rho}{2}\|C - (A - \text{diag}(A)\|_F^2 \end{aligned} \quad (32)$$

Denote by (C_k, A_k) the optimization variables at iteration k, by Δ_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 A_{k+1} by minimizing $\mathcal{L}$ with respect to A, while fixing (C_k, Δ_k). This is equivalent to solve the following problem:

$$\begin{aligned} &A_{k+1} = J - \text{diag}(J), \\ &J = \arg\min_{J} \frac{1}{2}\|C_k + \rho^{-1}\Delta_k - J\|_F^2 + \frac{\lambda}{\rho}\|J\|_{p,1}. \end{aligned} \quad (33)$$

(2) Obtain C_{k+1} by minimizing $\mathcal{L}$ with respect to C, while fixing (A_{k+1}, Δ_k). This is equivalent to solve the following problem:

$$\min_{C} \|X - XC\|_F^2 + \frac{\rho}{2}\|C - A_{k+1} + \frac{1}{\rho}\Delta_k\|_F^2 \quad (34)$$

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

$$C_{k+1} = (X^\top X + \frac{\rho}{2}I)^{-1}(X^\top X + \frac{\rho}{2}A_{k+1} - \frac{1}{2}\Delta_k). \quad (35)$$

(3) Obtain the Lagrangian multipliers (Δ_{k+1}) while fixing (C_{k+1}, A_{k+1}):

$$\Delta_{k+1} = \Delta_k + \rho(C_{k+1} - A_{k+1}). \quad (36)$$

(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