

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-representation 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:

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

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

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

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

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

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

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

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

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

whose solution for \mathbf{A} coincides with the solution of Eq. (1). By introducing a Lagrangian multipliers $\mathbf{\Delta}$ and a penalty parameter ρ , the Lagrangian function of the Eq. (3) can be written as

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

Denote by $(\mathbf{C}_k, \mathbf{A}_k)$ the optimization variables at iteration k , by $\mathbf{\Delta}_k$ the Lagrangian multipliers at iteration k , and by ρ_k the penalty parameter at iteration k . Taking derivatives 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 \mathbf{A}_{k+1} by minimizing \mathcal{L} with respect to \mathbf{A} , while fixing $(\mathbf{C}_k, \mathbf{\Delta}_k)$. This is equivalent to solve the fol-

lowing problem:

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

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

$$\begin{aligned} \mathbf{C}_{k+1} &= \arg \min_{\mathbf{C}} \frac{\rho_k}{2} \|\mathbf{E}_k - \mathbf{X} + \mathbf{X}\mathbf{C} + \rho_k^{-1} \Delta_k\|_F^2 \\ &+ \frac{\rho_k}{2} \|\mathbf{C} - \mathbf{A}_{k+1} + \rho_k^{-1} \delta_k\|_F^2 \end{aligned}$$

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

$$\begin{aligned} \mathbf{C}_{k+1} &= \mathbf{H} - \text{diag}(\mathbf{H}), \\ (\mathbf{X}^\top \mathbf{X} + \mathbf{I})\mathbf{H} &= \mathbf{X}^\top \mathbf{P}_k + \mathbf{Q}_k, \end{aligned} \quad (7)$$

where $\mathbf{P}_k = \mathbf{X} - \mathbf{E}_k - \rho_k^{-1} \Delta_k$ and $\mathbf{Q}_k = \mathbf{A}_{k+1} - \rho_k^{-1} \delta_k$.

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

$$\min_{\mathbf{E}} \frac{1}{2} \|(\mathbf{X} - \mathbf{X}\mathbf{C}_{k+1} - \rho_k^{-1} \Delta_k) - \mathbf{E}\|_F^2 + \rho_k^{-1} \|\mathbf{E}\|_1. \quad (8)$$

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

$$\mathbf{E}_{k+1} = \mathcal{S}_{\rho_k^{-1}}(\mathbf{X} - \mathbf{X}\mathbf{C}_{k+1} - \rho_k^{-1} \Delta_k), \quad (9)$$

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

(4) Obtain the Lagrangian multipliers $(\Delta_{k+1}, \delta_{k+1})$ while fixing $(\mathbf{C}_{k+1}, \mathbf{A}_{k+1}, \mathbf{E}_{k+1})$:

$$\begin{aligned} \Delta_{k+1} &= \Delta_k + \tau \rho_k (\mathbf{E}_{k+1} - \mathbf{X} + \mathbf{X}\mathbf{C}_{k+1}), \\ \delta_{k+1} &= \delta_k + \tau \rho_k (\mathbf{C}_{k+1} - \mathbf{A}_{k+1}), \end{aligned} \quad (10)$$

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 $\rho_{k+1} = \mu \rho_k$, where $\mu > 1$.

5 Non-Negative 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

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

We can also restrict that $\text{diag}(\mathbf{A}) = \mathbf{0}$ 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

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{X}\mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_{p,1}. \quad (12)$$

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

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

whose solution for \mathbf{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

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

Denote by $(\mathbf{C}_k, \mathbf{A}_k)$ the optimization variables at iteration k , by Δ_k the Lagrangian multipliers at iteration k , and by ρ_k the penalty parameter at iteration k . Taking derivatives 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 \mathbf{A}_{k+1} by minimizing \mathcal{L} with respect to \mathbf{A} , while fixing (\mathbf{C}_k, Δ_k) . This is equivalent to solve the following problem:

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{X}\mathbf{A}\|_F^2 + \frac{\rho}{2} \|\mathbf{A} - (\mathbf{C}_k + \rho_k^{-1} \Delta_k)\|_F^2, \quad (15)$$

which is equivalently to solve the following problem

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

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

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

and transform this problem as

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

which will save a lot of computational costs.

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

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

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

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

where \mathbf{F}_{i*} is the i th row of the matrix \mathbf{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 $(\mathbf{\Delta}_{k+1})$ while fixing $(\mathbf{C}_{k+1}, \mathbf{A}_{k+1})$:

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

(5) Update the penalty parameter ρ as $\rho_{k+1} = \mu \rho_k$, where $\mu > 1$.

7 Robust Large Scale Subset Selection via Dissimilarity based Outlier Detection

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

Then the proposed model can be formulated as

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

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

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

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

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

Denote by $(\mathbf{A}_k, \mathbf{C}_k)$ the optimization variables at iteration k , by $\mathbf{\Delta}_k$ the Lagrangian multiplier at iteration k , and by ρ_k the penalty parameter at iteration k . Taking derivatives 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 \mathbf{A}_{k+1} by minimizing \mathcal{L} with respect to \mathbf{A} , while fixing $(\mathbf{C}_k, \mathbf{\Delta}_k)$. This is equivalent to solve the following problem:

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

which is equivalently to solve the following problem

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

Since the matrices $\mathbf{X}^\top \mathbf{X}$ and $\mathbf{D}^\top \mathbf{D}$ are positive semi-definite 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:

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

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

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

where F_{i*} is the i th 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}) :

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

(5) Update the penalty parameter ρ as $\rho_{k+1} = \mu \rho_k$, 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_k the penalty parameter at iteration k . Taking derivatives 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:

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

(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:

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

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

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

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

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

(5) Update the penalty parameter ρ as $\rho_{k+1} = \mu \rho_k$, 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](#)