Unsupervised Person Re-identification Graph based LLE Model

1 Locally Linear Embedding

In vision and learning problems, we often have a set of data $X = \{x_1, \ldots, x_n\} \in \mathcal{R}^{d\times n}$ drawn from a union of c subspaces $\mathcal{S}_{s=1}^{c}$ where d is the feature dimension and n is the number of data vectors. To characterize the relation between the data in X the idea is to construct an affinity matrix $A \in \mathcal{R}^{n\times n}$ in which A_{ij} shows the affinity between X_i and X_j. While computing Euclidean distances on the raw data is the most intuitive way to construct the data affinity matrix, such metric usually does not reveal the global subspace structure of data well. Using the ideas from compressed sensing the general formulation is arrived at:

$$\begin{aligned}\min_{Z,E,W} \alpha||Z||_1 + \beta||E||_F + \gamma||W||_F \\ s.t.\ WX = WXZ + E\end{aligned} \tag{1}$$

where $Z \in \mathcal{R}^{n\times n}$ and $E \in \mathcal{R}^{d\times n}$ denote the representation matrix and the residual matrix, respectively. W acts as a metric.

2 Formulation

We also learn the affinity matrix by assuming that the features in X from same person should have larger probabilities to be in same cluster if their representations have smaller distance, and impose the following constraints,

$$\min_{A1=1,A\geq 0} ||z_i - zj||_F^2 a_{ij} \tag{2}$$

where $A \in \mathcal{R}^{n\times n}$ is the desired affinity matrix, whose element a_{ij} reflects the probability of the features x_i and x_j from the same cluster based on the distance between their representations z_i and z_j . The constraints $A1 = 1$ and $A \geq 0$ guarantee the probability property of each column of A. With some simple algebra, we integrate these constraints into (1), and have

$$\begin{aligned}\min_{Z,E,W} \alpha||Z||_1 + \beta||E||_F + \gamma tr(ZL_AZ^T) + \theta||W||_F + \lambda||A||_F^2 \\ s.t.\ WX = WXZ + E, A \geq 0, ||A|| = 1\end{aligned} \tag{3}$$

Where $L_A = D_A - A$ is the laplacian matrix of A and D_A is the degree matrix of A.

An auxiliary variable $Q \in \mathcal{R}^{d\times n}$ is used such that with the constraint that $Z = Q$ hence, it can be used to replace Z in the optimization expression. The constraints of the optimization problem are translated to an equivalent function which is the added to the expression. Hence the given equation becomes following equivalent Lagrangian

$$\mathcal{L}_{A1=1,A\geq 0}\begin{cases}\alpha||Z||_1 + \beta||E||_F + \gamma tr(QL_AQ^T) + \theta||W||_F + \lambda||A||_F^2 \\ +f(W, X, Z, E, A, Q) - \frac{1}{2\mu}\{||Y_1||_F - ||Y_2||_F\}\end{cases} \tag{4}$$

$$f(W,X,Z,E,A,Q) = \frac{\mu}{2}(||WX - WXZ - E + \frac{Y_1}{\mu}||_F^2 + ||Z - Q + \frac{Y_2}{\mu}||_F^2) \quad (5)$$

Where $\mu > 0$ which is a penalty parameter and the function $f(.)$ is defined as in equation (5) In the above equation, Y_1 and Y_2 are the Lagrangian multipliers. The ADMM alternatively updates one variable by minimizing $\mathcal{L}$ with fixing other variables. In addition to the Lagrangian multipliers, there are 5 variables, including Z, Q, E, A, W to be solved.

Here both Z and A give a notion of the cluster to which a sample belongs.

The solutions of these sub-problems are discussed below.

2.0.1 Solving for Z

With other variables in (5) fixed, the Z-sub problem can be written as:

$$min_Z \alpha ||Z||_1 + f(W,X,Z,E,A,Q) \quad (6)$$

to avoid the auxiliary variables and matrix inversions we used the linearized ADMM algorithm to minimize the z-subproblem. The quadratic term is replaced from the first order approximation at the previous iteration and adding a proximal term then z^{k+1} becomes:

$$arg \min_Z ||Z||_1 + \frac{\eta \mu^k}{2\alpha}||Z - Z^K||_F^2 + (\nabla_Z f^k, Z - Z^k) \quad (7)$$

where $\nabla_Z f^k = -\mu X^T W^T (WX - WXZ - E - \frac{y_1}{\mu})$

2.0.2 Solving for Q

By fixing rest of the variables we can obtain the optimization expression for the variable Q. Hence, Q-subproblem can be formulated as:

$$\min_Q \gamma tr(Q L_A Q^T) + ||Z - Q - \frac{y_2}{\mu}||_F^2 \quad (8)$$

To compute Q we differentiate the Lagrangian function with respect to Q and obtain the equation as: $Q^{k+1} = (Z^{K+1} + \frac{y_2^k}{\mu^k})(I + \gamma(L_{A^K} + L_{A^K}^T))^{-1}$

2.0.3 Solving for E

The E-subproblem can be formulated as follows when other variables are fixed:

$$E^{K+1} = arg \min_E ||E||_F^2 + \frac{\mu^k}{2\beta}||WX - WXZ^{k+1} - E - \frac{y_1^k}{\mu^k}||_F^2 \quad (9)$$

where differentiating with respect to E will give us considering the formulation in 4.

$$2\beta E - 2\mu(WX - WXZ^{k+1} - E + \frac{y_1^k}{\mu^k}) \quad (10)$$

2.0.4 Solving for W

The W-subproblem can be formulated as follows when other variables are fixed:

$$W^{K+1} = arg\min_{W} ||W||_F^2 + \frac{\mu^k}{2\theta}||WX - WXZ^{k+1} - E - \frac{y_1^k}{\mu^k}||_F^2 \tag{11}$$

where differentiating with respect to W will give us considering the formulation in 4.

$$2\theta W - 2\mu X(I - Z^{k+1})(WX(I - Z^{k+1}) - E + \frac{y_1}{\mu}) \tag{12}$$

2.0.5 Solving for A

THe subproblem for A can be formulated as:

$$\min_{A1=1, A\geq 0} \frac{\lambda}{2}||A||_F^2 + \gamma tr(QL_AQ^T) \tag{13}$$

We separate the above equation into a set of different and independent problems, and each a_i can be computed efficiently

$$a_i^{k+1} = (\frac{1 + \sum_{j=1}^{\epsilon}}{\epsilon}1 - u_i^{Q^{k+1}})_+ \tag{14}$$

where $u_i^{Q^{k+1}} \in \mathcal{R}^{n\times 1}$ is a vector whose j^{th} element is defined as $u_{ij}^{Q^{k+1}} = \frac{\frac{\gamma}{2}||q_i^{k+1} - q_j^{k+1}||_F^2}{\lambda}$.Also notice, that ϵ is a parameter to control the number of nearest neighbours to q_i.