Model

1 RPCA Problem

For a RPCA problem, given a data matrix $X \in \mathbf{R}^{p\times n}$, and transform it to a related feature matrix $F \in \mathbf{R}^{D\times n}$ find L and S that solve the problem:

$$\min_{L,S} \quad rank(L) + \|S\|_0$$

$$s.t. \quad F = L + S$$

Reformulate it as follows:

$$\min_{L,S} \quad \|L\|_* + \|S\|_1$$

$$s.t. \quad F = L + S$$

For an $m \times n$ matrix M with SVD $US'V^T$:

$$UT_\epsilon(S')V^T = \arg\min_X \quad \epsilon\|X\|_* + \frac{1}{2}\|X - M\|_F^2$$

$$T_\epsilon(M) = \arg\min_X \quad \epsilon\|X\|_1 + \frac{1}{2}\|X - M\|_F^2$$

where the $T_\epsilon(M)$ is the soft threshold operator:

$$T_\epsilon(M) = \begin{cases} M - \epsilon, & M > \epsilon \\ M + \epsilon, & M < -\epsilon \\ \quad 0 \quad , & otherwise \end{cases}$$

and S' is from $SVD(M) = US'V^T$. Consider the problem:

$$\min_X \quad f(X)$$

$$s.t. \quad c_j(X) = 0, j = 1, \cdots, m$$

By using the Augmented Lagrange Multipliers Method:
1.Initialize Λ, $\mu > 0$, $\rho \geq 0$
(Repeat until convergence:)
2.Compute $X = \arg\min_X \quad L(X)$ where

$$L(X) = f(X) + <\Lambda, C(X)> + \frac{\mu}{2}\|C(X)\|_F^2$$

3.Update $\Lambda = \Lambda + \mu C(X)$
4.Update $\mu = \rho\mu$

2 ALM-RPCA

With ALM, RPCA problem is reformulated as:

$$\min_{L,S} \quad \|L\|_* + \lambda\|S\|_1 + <Y, F - L - S> + \frac{\mu}{2}\|F - L - S\|_F^2$$

By using the Alternating Direction Method(ADM), we have:

2.1 Updating S with L fixed:

$$\begin{aligned}
&\min_S \quad \lambda\|S\|_1 + < Y, F-L-S > + \frac{\mu}{2}\|F-L-S\|_F^2 \\
&\min_S \quad \frac{\lambda}{\mu}\|S\|_1 + tr(\frac{Y^T}{\mu}(F-L-S)) + \frac{1}{2}\|F-L-S\|_F^2 + \frac{1}{2}\|\frac{Y}{\mu}\|_F^2 \\
&\min_S \quad \frac{\lambda}{\mu}\|S\|_1 + \frac{1}{2}\|S-(F-L+\frac{Y}{\mu})\|_F^2
\end{aligned}$$

So here we have the solution:

$$S = T_{\frac{\lambda}{\mu}(F-L+\frac{Y}{\mu})}$$

2.2 Updating L with S fixed:

$$\begin{aligned}
&\min_L \quad \|L\|_* + < Y, F-L-S > + \frac{\mu}{2}\|F-L-S\|_F^2 \\
&\min_L \quad \frac{1}{\mu}\|L\|_* + tr(\frac{Y^T}{\mu}(F-L-S)) + \frac{1}{2}\|F-L-S\|_F^2 + \frac{1}{2}\|\frac{Y}{\mu}\|_F^2 \\
&\min_L \quad \frac{1}{\mu}\|L\|_* + \frac{1}{2}\|L-(F-S+\frac{Y}{\mu})\|_F^2
\end{aligned}$$

So here we have the solution:

$$L = UT_{\frac{1}{\mu}}(S')V^T$$

where S' is from $SVD(F-S+\frac{Y}{\mu}) = US'V^T$. And every time we update

$$Y = Y + \mu(F-L-S)$$

$$\mu = \rho\mu$$

until convergence.Typical initialization:

1.$Y = \frac{sgn(F)}{max(\|F\|_2, \frac{\|F\|_\infty}{\lambda})}$, $\|F\|_2$ is spectral norm, largest singular value of elements of F, and $\|F\|_\infty$ is the largest absolute value of elements of F.
2.$\mu = 1.25\|F\|_2$.
3.$\rho = 1.5$.
4.$\lambda = 1/\sqrt{max(D,n)}$ for $D \times n$ matrix F.

3 About Our Model

Suppose that we have n samples $X = \{x_i\}_{i=1}^n \in \mathbf{R}^{p\times n}$. We aim to learn a set of binary codes $B = \{b_i\}_{i=1}^n \in \{-1,1\}^{L\times n}$ to well preserve their spatial structure, where the i^{th} column b_i is the L-bits codes for x_i. To take advantage of the label information, we're going to introduce the ground truth label matrix $Y = \{y_i\}_{i=1}^n \in \mathbf{R}^{C\times n}$. And at the same time we want to try a hash model with the idea we got from saliency detection. So the model can first written as:

$$\min_{W,B,L,S} \quad \|Y - W^TB\|_F^2 + \lambda_1\|B^TB - S^TS\|_F^2 + \lambda_2\|L\|_* + \lambda_3\|S\|_1$$

$$s.t. \;\; L+S=F, \;\; B\in\{-1,1\}^{L\times n}, \;\; B1_n = 0_L, \;\; BB^T = nI_L$$

The W is of a size $L\times C$, and W^T is of a size $C\times L$. Here we also have $y_i = [w_1^Tb_i, \cdots, w_C^Tb_i]^T$. w_k is the classification vector for class $k(k=1,\cdots,C)$ and $y_i \in \mathbf{R}^{C\times 1}$ is the label vector, of which the maximum item indicates the assigned class of x_i. F is the feature matrix and we want to decompose it as two parts with one low-rank part L and the other one sparse part S. The constraints here are balance constraint and decorrelation constraint.

4 Why Gradient Method Doesn't Work

Here we use ADM with gradient method, and here we don't consider the balance constraint and decorrelation constraint:

4.1 Updating W with others fixed:

Here the model can be written as:

$$\min_{W} \quad \|Y - W^T B\|_F^2$$

It is equivalent to solve the optimization problem:

$$\min_{W} \quad \frac{1}{2} tr[(B^T W - Y^T)^T (B^T W - Y^T)]$$

it is the least squares problem, the solution is:

$$W = (BB^T)^{-1} BY^T$$

4.2 Updating B with others fixed:

In this part, the model:

$$\min_{B} \quad H(B) = \|Y - W^T B\|_F^2 + \lambda_1 \|B^T B - S^T S\|_F^2$$

So the gradient is:

$$\nabla_B H(B) = WW^T B - WY + 4\lambda_1 B^T (B^T B - S^T S)$$

So the biggest problem is how can I solve B from $\nabla_B H(B) = 0$, it is a cubic equation with the variable a matrix.

5 Introduce Auxiliary Variables

Here I want to rewrite the model as:

$$\min_{W,B,Z,L,S,M,H} \quad \|Y - W^T B\|_F^2 + \lambda_1 \|B^T Z - M^T H\|_F^2 + \alpha \|B - Z\|_F^2 + \lambda_2 \|L\|_* + \lambda_3 \|S\|_1$$

$$s.t. \begin{cases} B \in \{-1, +1\}^{L \times n}, \\ Z \in R^{L \times n}, Z 1_n = 0_L, ZZ^T = nI_L \\ F = L + S, S = M, M = H \end{cases}$$

With the constraints, the model can be rewritten as:

$$\min_{W,B,Z,L,S,M,H} \begin{cases} \|Y - W^T B\|_F^2 + \lambda_1 \|B^T Z - M^T H\|_F^2 + \alpha \|B - Z\|_F^2 \\ + \lambda_2 \|L\|_* + \lambda_3 \|S\|_1 + < Y_1, F - L - S > + \frac{\mu}{2} \|F - L - S\|_F^2 \\ + < Y_2, M - S > + \frac{\mu}{2} \|M - S\|_F^2 + < Y_3, M - H > + \frac{\mu}{2} \|M - H\|_F^2 \end{cases}$$

5.1 Updating W with others fixed:

The updating process is:

$$\min_{W} \quad \|Y - W^T B\|_F^2$$
$$\min_{W} \quad \|B^T W - Y^T\|_F^2$$

The solution is

$$W = (BB^T)^{-1} BY^T$$

5.2 Updating B with others fixed:

The updating process is:

$$\begin{aligned}&\min_{B} \quad \|Y - W^T B\|_F^2 + \lambda_1 \|B^T Z - M^T H\| + \alpha \|B - Z\|_F^2 \\ &s.t. \quad B \in \{-1, +1\}^{L \times n}\end{aligned}$$

This is equivalent to the optimization problem:

$$\begin{aligned}&\max_{B} \quad tr(B^T(WY + \lambda_1 Z H^T M + \alpha Z)) \\ &s.t. \quad B \in \{-1, +1\}^{L \times n}\end{aligned}$$

It has a closed form solution:

$$B = sgn(WY + \lambda_1 Z H^T M + \alpha Z)$$

5.3 Updating Z with others fixed:

The updating process is:

$$\begin{aligned}&\min_{Z} \quad \lambda_1 \|B^T Z - M^T H\|_F^2 + \alpha \|B - Z\|_F^2 \\ &s.t. \quad Z \in \mathbf{R}^{L \times n}, Z 1_n = 0_L, Z Z^T = n I_L\end{aligned}$$

It can be further reduced to:

$$\begin{aligned}&\max_{Z} \quad tr(Z^T(\lambda_1 B M^T H + \alpha B)) \\ &s.t. \quad Z \in \mathbf{R}^{L \times n}, Z 1_n = 0_L, Z Z^T = n I_L\end{aligned}$$

Here we introduce two variables E and J, let $E = \lambda_1 B M^T H + \alpha B$ and $J = I_N - \frac{1}{N} 1_n 1_n^T$:

$$\begin{aligned}&J E^T = U \Sigma V^T = \Sigma_{k=1}^{K'} \sigma_k u_k v_k^T \\ &U = [u_1, u_2, \cdots, u_{K'}] \quad and \quad V = [v_1, v_2, \cdots, v_{K'}]\end{aligned}$$

Note that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{K'} > 0$. Then, by employing a Gram-Schmidt process one can easily construct matrices

$$\bar{U} \in \mathbf{R}^{n \times (L-K')} \quad and \quad \bar{V} \in \mathbf{R}^{L \times (L-K')}$$

$$s.t. \quad \begin{cases} \bar{U}^T \bar{U} = I_{L-K'}, \ [U \ 1_n]^T \bar{U} = 0_{(K'+1) \times (L-K')} \\ \bar{V}^T \bar{V} = I_{L-K'}, \ V^T \bar{V} = 0_{K' \times (L-K')} \end{cases}$$

Here

$$E J E^T = [V \ \bar{V}] \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} [V \ \bar{V}]^T$$

so we get the V, $\bar{V}$ and Σ, and then immediately leads to $U = J E^T V \Sigma^{-1}$. The matrix $\bar{U}$ is set to a random matrix followed by Gram-Schmidt process. It can be seen that Z is uniquely optimal when L=K', which means $J E^T$ is full column rank. (Actually I guess $J E^T$ is full column rank in high probability.) So we have a closed form for Z:

$$Z = \sqrt{N} [V \ \bar{V}] [U \ \bar{U}]^T$$

5.4 Updating L with others fixed:

The process is:

$$\begin{aligned}&\min_{L} \quad \lambda_2 \|L\|_* + < Y_1, F - L - S > + \frac{\mu}{2} \|F - L - S\|_F^2 \\ &\min_{L} \quad \frac{\lambda_2}{\mu} \|L\|_* + \frac{1}{2} \|L - (F - S + \frac{Y_1}{\mu})\|_F^2\end{aligned}$$

The solution is:

$$L = U T_{\frac{\lambda_2}{\mu}}(S') V^T$$

where $U S' V^T = SVD(F - S + \frac{Y_1}{\mu})$.

5.5 Updating S with others fixed:

The process is:

$$\min_{S} \quad \lambda_3\|S\|_1 + < Y_1, F-L-S > + \frac{\mu}{2}\|F-L-S\|_F^2 + < Y_2, M-S > + \frac{\mu}{2}\|M-S\|_F^2$$
$$\min_{S} \quad \frac{\lambda_3}{\mu}\|S\|_1 + \frac{1}{2}\|S-(F-L+\frac{Y_1}{\mu})\|_F^2 + \frac{1}{2}\|S-(M+\frac{Y_1}{\mu})\|_F^2$$

So the solution is:

$$S = T_{\frac{\lambda_3}{\mu}}(F-L+M+2\frac{Y_1}{\mu})$$

5.6 Updating M with others fixed:

The process is:

$$\min_{M} \quad \lambda_1\|H^T M - Z^T B\|_F^2 + < Y_2, M-S > + \frac{\mu}{2}\|M-S\|_F^2 + < Y_3, M-H > + \frac{\mu}{2}\|M-H\|_F^2$$
$$\min_{M} \quad \lambda_1\|H^T M - Z^T B\|_F^2 + \frac{\mu}{2}(\|M-S+\frac{Y_2}{\mu}\|_F^2 + \|M-H+\frac{Y_3}{\mu}\|_F^2)$$

The solution is(Steps refer to Z-stage):

$$M = \sqrt{N}[V' \ \bar{V}'][U' \ \bar{U}']^T$$

5.7 Updating H with others fixed:

The process:

$$\min_{H} \quad \lambda_1\|H^T M - Z^T B\|_F^2 + < Y_3, M-H > + \frac{\mu}{2}\|M-H\|_F^2$$
$$\min_{H} \quad \lambda_1\|H^T M - Z^T B\|_F^2 + \frac{\mu}{2}\|M-H+\frac{Y_3}{\mu}\|_F^2$$

Next steps refer to Z-stage.

$$H = \sqrt{N}[V'' \ \bar{V}''][U'' \ \bar{U}'']^T$$

6 If We Need the Relation Between X and B Matrix

7 How to Construct F Matrix