Data Mining Problem

1 Problem Setting

$$
\begin{aligned}
\min_{W} L = {} & \lambda_0 \sum_{k=1}^{K} \sum_{n=1}^{N} (X_n^k {W_{in}}_n^k - {Y_{in}}_n^k)^2 + \frac{1}{2}\lambda_1 \sum_{k=1}^{K} tr\left({W_{in}}^k L_{sp} ({W_{in}}^k)^T\right) + \sum_{n=1}^{N} \lambda_2 tr\left({W_{in}}_n L_{te} ({W_{in}}_n)^T\right) \\
& + \lambda_3 \sum_{k=1}^{K} \sum_{n=1}^{N} (X_n^k {W_{out}}_n^k - {Y_{out}}_n^k)^2 + \lambda_4 \frac{1}{2} \sum_{k=1}^{K} tr\left({W_{out}}^k L_{sp} ({W_{out}}^k)^T\right) + \lambda_5 \sum_{n=1}^{N} tr\left({W_{out}}_n L_{te} ({W_{out}}_n)^T\right) \\
& + \lambda_6 \sum_{k=1}^{K} \sum_{m=1}^{M} h(m) |\sum_{i \in C_m} X_i^k ({W_{in}}_i^k - {W_{out}}_i^k)| + \theta \left(\sum_{k=1}^{K} \sum_{n=1}^{N} (\|{W_{in}}_n^k\|^2) + \sum_{k=1}^{K} \sum_{n=1}^{N} (\|{W_{out}}_n^k\|^2) \right)
\end{aligned}
\tag{1.1}
$$

2 Optimization Method

We introduce auxiliary variable matrices:

$$
\begin{aligned}
E^k &= {W_{in}}^k L_{sp} ({W_{in}}^k)^T, \quad k = 1, \ldots, K \\
F_n &= {W_{in}}_n L_{temp} ({W_{in}}_n)^T, \quad n = 1, \ldots, N \\
G^k &= {W_{out}}^k L_{sp} ({W_{out}}^k)^T, \quad k = 1, \ldots, K \\
H_n &= {W_{out}}_n L_{temp} ({W_{out}}_n)^T, \quad n = 1, \ldots, N
\end{aligned}
$$

We can rewrite the original problem into

$$
\begin{aligned}
\min_{W} L = \lambda_0 & \sum_{k=1}^{K}\sum_{n=1}^{N}(X_n^k {W_{in}}_n^k - {Y_{in}}_n^k)^2 + \frac{1}{2}\lambda_1 \sum_{k=1}^{K} tr(E^k) + \lambda_2 \sum_{n=1}^{N} tr(F_n) \\
& + \lambda_3 \sum_{k=1}^{K}\sum_{n=1}^{N}(X_n^k {W_{out}}_n^k - {Y_{out}}_n^k)^2 + \frac{1}{2}\lambda_4 \sum_{k=1}^{K} tr(G^k) + \lambda_5 \sum_{n=1}^{N} tr(H_n) \\
& + \lambda_6 \sum_{k=1}^{K}\sum_{m=1}^{M}\left|\sum_{i\in C_m} X_i^k({W_{in}}_i^k - {W_{out}}_i^k)\right| + \theta\left(\sum_{k=1}^{K}\sum_{n=1}^{N}(\|{W_{in}}_n^k\|^2) + \sum_{k=1}^{K}\sum_{n=1}^{N}(\|{W_{out}}_n^k\|^2)\right) \\
\text{s.t. } & E^k = {W_{in}}^k L_{sp}({W_{in}}^k)^T, \quad k = 1, \dots, K \\
& F_n = {W_{in}}_n L_{temp}({W_{in}}_n)^T, \quad n = 1, \dots, N \\
& G^k = {W_{out}}^k L_{sp}({W_{out}}^k)^T, \quad k = 1, \dots, K \\
& H_n = {W_{out}}_n L_{temp}({W_{out}}_n)^T, \quad n = 1, \dots, N
\end{aligned}
\tag{2.1}
$$

$\mathbf{W}^k \in \mathbb{R}^{M\times N}$
$\mathbf{P} \in \mathbb{R}^{N\times N^2}$
$E^k \in \mathbb{R}^{M\times M}$
$U^k \in \mathbb{R}^{M\times M}$
$G^k \in \mathbb{R}^{M\times M}$
$G^k \in \mathbb{R}^{M\times M}$
$\mathbf{W}_n \in \mathbb{R}^{M\times(D*24)}$
$\mathbf{Q} \in \mathbb{R}^{(D*24)\times((D-1)*24)}$
$F_n \in \mathbb{R}^{M\times M}$
$V_n \in \mathbb{R}^{M\times M}$
$H_n \in \mathbb{R}^{M\times M}$
$R_n \in \mathbb{R}^{M\times M}$

The ADMM objective function of Eq(2.1) can be written as

$$
\begin{aligned}
\min\ & L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R) \\
= \lambda_0 & \sum_{k=1}^{K} \sum_{n=1}^{N} (X_n^k {W_{in}}_n^k - {Y_{in}}_n^k)^2 \\
& + \frac{1}{2}\lambda_1 \sum_{k=1}^{K} tr(E^k) + \frac{\rho}{2}\lambda_1 \sum_{k=1}^{K} \| {W_{in}}^k L_{sp} ({W_{in}}^k)^T - E^k + U^k \|_F^2 \\
& + \lambda_2 \sum_{n=1}^{N} tr(F_n) + \frac{\rho}{2}\lambda_2 \sum_{n=1}^{N} \| W_{inn} L_{temp} (W_{inn})^T - F_n + V_n \|_F^2 \\
& + \lambda_3 \sum_{k=1}^{K} \sum_{n=1}^{N} (X_n^k {W_{out}}_n^k - {Y_{out}}_n^k)^2 \\
& + \frac{1}{2}\lambda_4 \sum_{k=1}^{K} tr(G^k) + \frac{\rho}{2}\lambda_4 \sum_{k=1}^{K} \| {W_{out}}^k L_{sp} ({W_{out}}^k)^T - G^k + S^k \|_F^2 \\
& + \lambda_5 \sum_{n=1}^{N} tr(H_n) + \frac{\rho}{2}\lambda_5 \sum_{n=1}^{N} \| W_{outn} L_{temp} (W_{outn})^T - H_n + R_n \|_F^2 \\
& + \lambda_6 \sum_{k=1}^{K} \sum_{m=1}^{M} h(m) \left| \sum_{i \in C_m} X_i^k ({W_{in}}_i^k - {W_{out}}_i^k) \right| + \theta \left(\sum_{k=1}^{K} \sum_{n=1}^{N} (\| {W_{in}}_n^k \|^2) + \sum_{k=1}^{K} \sum_{n=1}^{N} (\| {W_{out}}_n^k \|^2) \right)
\end{aligned}
\tag{2.2}
$$

The $(t+1)$-th iteration of ADMM of (2.2) consists of the following procedures:

$$
\begin{aligned}
{W_{in}}_n^k(t+1) &= \underset{{W_{in}}_n^k}{\operatorname{argmin}} L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R) && (2.3)\\
E^k(t+1) &= \underset{E^k}{\operatorname{argmin}} \frac{1}{2} tr(E^k) + \frac{\rho}{2} \|{W_{in}}^k L_{sp} ({W_{in}}^k)^T - E^k + U^k\|_F^2 && (2.4)\\
&= -\frac{1}{2\rho} I + {W_{in}}^k L_{sp} ({W_{in}}^k)^T + U^k && (2.5)\\
F_n(t+1) &= \underset{F_n}{\operatorname{argmin}} tr(F_n) + \frac{\rho}{2} \|W_{inn} L_{te} (W_{inn})^T - F_n + V_n\|_F^2 && (2.6)\\
&= -\frac{1}{\rho} I + W_{inn} L_{te} (W_{inn})^T + V_n && (2.7)\\
{W_{out}}_n^k(t+1) &= \underset{{W_{out}}_n^k}{\operatorname{argmin}} L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R) && (2.8)\\
G^k(t+1) &= \underset{G^k}{\operatorname{argmin}} \frac{1}{2} tr(G^k) + \frac{\rho}{2} \|{W_{out}}^k L_{sp} ({W_{out}}^k)^T - G^k + S^k\|_F^2 && (2.9)\\
&= -\frac{1}{2\rho} I + {W_{out}}^k L_{sp} ({W_{out}}^k)^T + S^k && (2.10)\\
H_n(t+1) &= \underset{H_n}{\operatorname{argmin}} tr(H_n) + \frac{\rho}{2} \sum_{n=1}^{N} \|W_{outn} L_{te} (W_{outn})^T - H_n + R_n\|_F^2 && (2.11)\\
&= -\frac{1}{\rho} I + W_{outn} L_{te} (W_{outn})^T + R_n && (2.12)\\
U^k(t+1) &= U^k + {W_{in}}^k L_{sp} ({W_{in}}^k)^T - E^k && (2.13)\\
V_n(t+1) &= V_n + W_{inn} L_{te} (W_{inn})^T - F_n && (2.14)\\
S^k(t+1) &= S^k + {W_{out}}^k L_{sp} ({W_{out}}^k)^T - G^k && (2.15)\\
R_n(t+1) &= R_n + W_{outn} L_{te} (W_{outn})^T - H_n && (2.16)
\end{aligned}
$$

In the update of ${W_{in}}_n^k$ and ${W_{out}}_n^k$, we use SGD to get the update formula. The gradient of $L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)$ with respect to ${W_{in}}_n^k$ is :

$$
\begin{aligned}
&\frac{\partial L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)}{\partial {W_{in}}_n^k}\\
&= 2(X_n^k {W_{in}}_n^k - {Y_{in}}_n^k) \cdot (X_n^k)^T\\
&\quad + \rho \left({W_{in}}^k L_{sp} ({W_{in}}^k)^T - (E^k(t))^T + U^k(t) \right) \cdot {W_{in}}^k \cdot L_{sp}\\
&\quad + \rho \left({W_{in}}^k L_{sp} ({W_{in}}^k)^T - E^k(t) + U^k(t) \right) \cdot {W_{in}}^k \cdot L_{sp}\\
&\quad + \rho \left(W_{inn} L_{te} (W_{inn})^T - (F^n(t))^T + V^n(t) \right) \cdot W_{inn} \cdot L_{te}\\
&\quad + \rho \left(W_{inn} L_{te} (W_{inn})^T - F^n(t) + V^n(t) \right) \cdot W_{inn} \cdot L_{te}\\
&\quad + 2\theta {W_{in}}_n^k + \frac{\partial g(W)}{\partial {W_{in}}_n^k}
\end{aligned} \qquad (2.17)
$$

where $g(W) = \sum_{k=1}^{K} \sum_{m=1}^{M} h(m) \left| \sum_{i \in C_m} X_i^k ({W_{in}}_i^k - {W_{out}}_i^k) \right|$ and

$$\frac{\partial g(W)}{\partial {W_{in}}_n^k} = \text{sgn}\left(\sum_{i \in C_m} X_i^k ({W_{in}}_i^k - {W_{out}}_i^k) \right) \cdot (X_n^k)^T \tag{2.18}$$

where

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

Algorithm 1 The ADMM Framework

Input:
Output:

Initialize
while Not Convergent **do**
  Select integer $k \in [1, K]$ and $n \in [1, N]$ randomly
  while Not Convergent **do**
    Calculate $\frac{\partial L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)}{\partial {W_{in}}_n^k}$ according to Eq.(2.14)
    Update ${W_{in}}_n^k \leftarrow {W_{in}}_n^k - \gamma \frac{\partial L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)}{\partial {W_{in}}_n^k}$
  end while
  Update E^k according to Eq.(2.4)
  Update F_n according to Eq.(2.5)
  while Not Convergent **do**
    Calculate $\frac{\partial L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)}{\partial {W_{out}}_n^k}$
    Update ${W_{out}}_n^k \leftarrow {W_{out}}_n^k - \gamma \frac{\partial L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)}{\partial {W_{out}}_n^k}$
  end while
  Update G^k according to Eq.(2.7)
  Update H_n according to Eq.(2.8)
  Update U^k according to Eq.(2.9)
  Update V_n according to Eq.(2.10)
  Update S^k according to Eq.(2.11)
  Update R_n according to Eq.(2.12)
end while