Appendix of "Bi-level Ensemble Method for Unsupervised Feature Selection"

Appendix A: Proof of Theorem 1

Theorem 1. *Denoting* $B_{ij} = \|\mathbf{P}diag(\mathbf{v})\mathbf{x}_i - \mathbf{P}diag(\mathbf{v})\mathbf{x}_j\|_2^2$ *and* $C_{ij} = \|\mathbf{Y}_{i.} - \mathbf{Y}_{j.}\|_2^2$, *the closed-form solution of the subproblem w.r.t.* $\mathbf{S}$ *is*

$$S_{ij} = \max\left(\min\left(\frac{\sum_{k=1}^{m}\beta_k^2 S_{ij}^{(k)} - \frac{B_{ij}+\rho C_{ij}}{2W_{ij}^2}}{\sum_{k=1}^{m}\beta_k^2}, 1\right), 0\right). \tag{1}$$

Proof. When fixing other variables, by taking the definition of B_{ij} and C_{ij} into the objective function, we reformulate it as follows:

$$\begin{aligned}&\min_{\mathbf{S}} \sum_{i,j=1}^{n} B_{ij}S_{ij} + \sum_{k=1}^{m}\beta_k^2\|\mathbf{W}\odot(\mathbf{S}-\mathbf{S}^{(k)})\|_F^2 + \rho\sum_{i,j=1}^{n} C_{ij}S_{ij},\\ &s.t. 0 \le S_{ij} \le 1, \quad \mathbf{S} = \mathbf{S}^T.\end{aligned} \tag{2}$$

For simplicity, we first remove the constraint $\mathbf{S} = \mathbf{S}^T$, and at last we show the learned $\mathbf{S}$ satisfies the symmetric constraint. When removing the symmetric constraint, we can decouple Eq. (2) into $n \times n$ independent subproblems. Consider the (i, j)-th subproblem:

$$\begin{aligned}\min_{S_{ij}} \quad & \sum_{k=1}^{m}\beta_k^2 W_{ij}^2(S_{ij} - S_{ij}^{(k)})^2 + (B_{ij} + \rho C_{ij})S_{ij},\\ s.t. \quad & 0 \le S_{ij} \le 1.\end{aligned} \tag{3}$$

By setting its derivation w.r.t. S_{ij} to zero, we obtain

$$S_{ij} = \frac{\sum_{k=1}^{m}\beta_k^2 S_{ij}^{(k)} - \frac{B_{ij}+\rho C_{ij}}{2W_{ij}^2}}{\sum_{k=1}^{m}\beta_k^2}. \tag{4}$$

If $\frac{\sum_{k=1}^{m}\beta_k^2 S_{ij}^{(k)} - \frac{B_{ij}+\rho C_{ij}}{2W_{ij}^2}}{\sum_{k=1}^{m}\beta_k^2} > 1$, Eq. (3) is a monotone decreasing function in the range $[0, 1]$, and thus the solution is 1. Similarly, if $\frac{\sum_{k=1}^{m}\beta_k^2 S_{ij}^{(k)} - \frac{B_{ij}+\rho C_{ij}}{2W_{ij}^2}}{\sum_{k=1}^{m}\beta_k^2} < 0$, the optima is

0. Therefore, the solution of $\mathbf{S}$ is

$$S_{ij} = \max\left(\min\left(\frac{\sum_{k=1}^{m}\beta_k^2 S_{ij}^{(k)} - \frac{B_{ij}+\rho C_{ij}}{2W_{ij}^2}}{\sum_{k=1}^{m}\beta_k^2}, 1\right), 0\right).$$

Now, we prove that it is symmetric. we use the mathematical induction. In the first iteration, we initialize $\mathbf{S} = \frac{1}{m}\sum_{k=1}^{m}\mathbf{S}^{(k)}$, which is symmetric. Then, in the following iterations, if $\mathbf{S}$ in the last iteration is symmetric, then according to the optimization of $\mathbf{W}$, $\mathbf{W}$ will be symmetric too. Moreover, it is easy to verify that $\mathbf{B}$ and $\mathbf{C}$ are always symmetric, and thus $\mathbf{S}$ obtained by Eq. (1) satisfies the symmetric constraint. □

Appendix B: Proof of Theorem 2

Theorem 2. *The following subproblem of* $\mathbf{v}$ *is strictly convex quadratic programming.*

$$\begin{aligned}
&\min_{\mathbf{v}} \sum_{i,j=1}^{n} \|\mathbf{P}diag(\mathbf{v})\mathbf{x}_i - \mathbf{P}diag(\mathbf{v})\mathbf{x}_j\|_2^2 S_{ij} + \sum_{k=1}^{m}\alpha_k^2\|\mathbf{v}-\mathbf{v}^{(k)}\|_2^2 \\
&s.t. \quad 0 \le v_i \le 1, \quad \sum_{i=1}^{d} v_i = 1.
\end{aligned} \tag{5}$$

Proof. Obviously, Eq. (5) is a quadratic programming, and we just need to prove its convexity. The constraint is a convex set. So, we focus on the objective function. We reformulate Eq. (5) as follows:

$$\begin{aligned}
&\min_{\mathbf{v}} \quad \mathbf{v}^T\left((\mathbf{XLX}^T)\odot(\mathbf{P}^T\mathbf{P}) + \sum_{k=1}^{m}\alpha_k^2\mathbf{I}\right)\mathbf{v} - \mathbf{v}^T\sum_{k=1}^{m}\alpha_k^2\mathbf{v}^{(k)}, \\
&s.t. \qquad 0 \le v_i \le 1, \quad \sum_{i=1}^{d} v_i = 1,
\end{aligned} \tag{6}$$

where $\mathbf{I}$ is an Identity matrix.

To prove its convexity, we just need to prove $(\mathbf{XLX}^T)\odot(\mathbf{P}^T\mathbf{P}) + \sum_{k=1}^{m}\alpha_k^2\mathbf{I}$ is a positive semi-definite (P.S.D.) matrix. Since $\mathbf{L}$ is a Laplacian matrix, it is a P.S.D. matrix, and $\mathbf{XLX}^T$ is also P.S.D. It is easy to verify that $\mathbf{P}^T\mathbf{P}$ is also P.S.D. Then we show that the Hadamard product of two P.S.D. matrix is also P.S.D.

Lemma 1. *If both* $\mathbf{M}$ *and* $\mathbf{N}$ *are P.S.D. matrices, then* $\mathbf{M}\odot\mathbf{N}$ *is also a P.S.D. matrix.*

Proof. If $\mathbf{N}$ is P.S.D., we can find a matrix $\mathbf{Q}$ such that $\mathbf{N} = \mathbf{Q}^T\mathbf{Q}$. Then, given any vector $\mathbf{x}$, we compute $\mathbf{x}^T(\mathbf{M}\odot\mathbf{N})\mathbf{x}$ as follows:

$$
\begin{aligned}
\mathbf{x}^T(\mathbf{M}\odot\mathbf{N})\mathbf{x} =& \sum_{i,j} x_i M_{ij} N_{ij} x_j \\
=& \sum_{i,j} x_i M_{ij} \sum_k Q_{ki} Q_{kj} x_j \\
=& \sum_k (\mathbf{Q}_{k.}\odot\mathbf{x})^T \mathbf{M} (\mathbf{Q}_{k.}\odot\mathbf{x}) \\
\geq& 0
\end{aligned} \tag{7}
$$

The last inequality is due to that $\mathbf{M}$ is P.S.D. For any vector $\mathbf{x}$, we have $\mathbf{x}^T(\mathbf{M}\odot\mathbf{N})\mathbf{x} \geq 0$, and thus $\mathbf{M}\odot\mathbf{N}$ is also a P.S.D. matrix. □

According to Lemma 1, we have $(\mathbf{X}\mathbf{L}\mathbf{X}^T)\odot(\mathbf{P}^T\mathbf{P})$ is a P.S.D. matrix. Since $\sum_{k=1}^m \alpha_k^2 > 0$, $(\mathbf{X}\mathbf{L}\mathbf{X}^T)\odot(\mathbf{P}^T\mathbf{P})+\sum_{k=1}^m \alpha_k^2\mathbf{I}$ is a positive definite matrix. Therefore, Eq. (5) is strictly convex quadratic programming. □