Supplementary Material for GSPL: A Succinct Kernel Model for Group-Sparse Projections Learning of Multiview Data

Danyang Wu[1], Jin Xu[2], Xia Dong[1], Meng Liao[2], Rong Wang[3], Feiping Nie[1], and Xuelong Li[1]

[1]School of Computer Science and School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China
[2]Data Quality Team, WeChat, Tencent Inc., Guangdong, P. R. China
[3]School of Cybersecurity and School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China

danyangwu41x@mail.nwpu.edu.cn, {jinxxu, maricoliao}@tencent.com, {xiadongpgh, feipingnie}@gmail.com, wangrong07@tsinghua.org.cn, li@nwpu.edu.cn

Abstract

The supplementary material is attached here to complete the article entitled "GSPL: A Succinct Kernel Model for Group-Sparse Projections Learning of Multiview Data". The following aspects are presented here: the solution of problem (16) and the proof of Remark 1.

1 Solution of Problem (16)

In this section, we give the details in terms of the solution of problem (16). Problem (16) in this article corresponds to the problem (2) here. We first start with problem (15) in this article corresponding to the problem (1) here.

$$\max_{\widehat{\mathbf{W}}^{\mathrm{T}}\widehat{\mathbf{W}}=\mathbf{I}_{\mathrm{m}\times\mathrm{m}},\ \|\widehat{\mathbf{W}}\|_{2,0}=\mathrm{k}} \mathbf{Tr}\left(\widehat{\mathbf{W}}^{\mathrm{T}}\mathbf{S}_{\mathbf{E}}\widehat{\mathbf{W}}\right), \tag{1}$$

which is NP-hard, then we consider to solve it into two cases. At first, we consider the case $\mathbf{rank}(\mathbf{S}_{\mathbf{E}}) \leqslant \mathrm{m}$. Since $\|\widehat{\mathbf{W}}\|_{2,0} = \mathrm{k}$, suppose [1]$\mathbf{q} \in \mathbf{Ind}(\mathrm{k}, \widehat{\mathrm{d}})$ is the $\mathbb{R}^{\mathrm{k}}$ indicator vector of non-sparse rows of $\widehat{\mathbf{W}}$, then $\widehat{\mathbf{W}}$ can be decomposed into $\widehat{\mathbf{W}} = \mathbf{BD}$. Wherein, $\mathbf{B} = \mathbf{\Pi}(\mathbf{q}) \in \{\mathbf{0}, \mathbf{1}\}_{\widehat{\mathrm{d}}\times\mathrm{k}}$, whose $\langle \mathrm{u}, \mathrm{v}\rangle$-th element $\mathrm{b}_{\mathrm{uv}} = 1$ only if $\mathrm{u} = \mathrm{q}_{\mathrm{v}}$, and $\mathbf{\Pi}(\cdot)$ is the mapping function $\mathbf{Ind}(\mathrm{k}, \widehat{\mathrm{d}}) \to \{\mathbf{0}, \mathbf{1}\}_{\widehat{\mathrm{d}}\times\mathrm{k}}$; $\mathbf{D} \in \mathbb{R}^{\mathrm{k}\times\mathrm{m}}$ and the u-th row of $\mathbf{D}$ is the q_{u}-th row of $\widehat{\mathbf{W}}$, and naturally $\mathbf{D}^{\mathrm{T}}\mathbf{D} = \mathbf{I}_{\mathrm{m}\times\mathrm{m}}$. Then problem (1) can be written as the following problem *w.r.t.* $\mathbf{B}$ and $\mathbf{D}$:

$$\max_{\mathbf{B}=\mathbf{\Pi}(\mathbf{q}),\mathbf{q}\in\mathbf{Ind}(\mathrm{k},\widehat{\mathrm{d}}),\mathbf{D}^{\mathrm{T}}\mathbf{D}=\mathbf{I}_{\mathrm{m}\times\mathrm{m}}} \mathbf{Tr}\left(\mathbf{D}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{S}_{\mathbf{E}}\mathbf{B}\mathbf{D}\right). \tag{2}$$

[1]$\mathbf{Ind}(\mathrm{k}, \widehat{\mathrm{d}})$ is a $\mathbb{R}^{\mathrm{k}}$ indicator vector, which selectes k elements from $\{1, \ldots, \widehat{\mathrm{d}}\}$ as ascending order and the elements are not duplicated. For example, $\mathbf{Ind}(2, 3)$ can be $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$.

According to Ky Fan's theorem, problem (2) *w.r.t.* $\mathbf{B}$ can be written as

$$\max_{\mathbf{B}=\mathbf{\Pi}(\mathbf{q}),\ \mathbf{q}\in\mathbf{Ind}(\mathrm{k},\widehat{\mathrm{d}})} \sum_{\mathrm{t}=1}^{\mathrm{m}} \lambda_{\mathrm{t}}\left(\mathbf{B}^{\mathrm{T}}\mathbf{S}_{\mathbf{E}}\mathbf{B}\right), \tag{3}$$

where $\lambda_{\mathrm{t}}(\cdot)$ is the t-th eigenvalue of $(\cdot)$. Considering $\mathbf{rank}(\mathbf{S}_{\mathbf{E}}) \leqslant \mathrm{m} \to \mathbf{rank}(\mathbf{B}^{\mathrm{T}}\mathbf{S}_{\mathbf{E}}\mathbf{B}) \leqslant \mathrm{m}$ and $\mathrm{m} \leqslant \mathrm{k} \leqslant \mathrm{d}$, we have

$$\sum_{\mathrm{t}=1}^{\mathrm{m}} \lambda_{\mathrm{t}}\left(\mathbf{B}^{\mathrm{T}}\mathbf{S}_{\mathbf{E}}\mathbf{B}\right) = \mathbf{Tr}\left(\mathbf{B}^{\mathrm{T}}\mathbf{S}_{\mathbf{E}}\mathbf{B}\right). \tag{4}$$

Then problem (3) can be written as

$$\max_{\mathbf{B}=\mathbf{\Pi}(\mathbf{q}),\ \mathbf{q}\in\mathbf{Ind}(\mathrm{k},\widehat{\mathrm{d}})} \mathbf{Tr}\left(\mathbf{B}^{\mathrm{T}}\mathbf{S}_{\mathbf{E}}\mathbf{B}\right), \tag{5}$$

where the optimal $\mathbf{B} = \mathbf{\Pi}(\tilde{\mathbf{q}})$, where $\tilde{\mathbf{q}} \in \mathbf{Ind}(\mathrm{k}, \widehat{\mathrm{d}})$ is the indicate vector of the first k largest values of the diagonal vector of $\mathbf{S}_{\mathbf{E}}$. Then in problem (2), the solution $\mathbf{D}$ can be formed by the eigenvectors corresponding to first-m largest eigenvalues of $\mathbf{B}^{\mathrm{T}}\mathbf{S}_{\mathbf{E}}\mathbf{B}$. Finally, $\widehat{\mathbf{W}}$ can be calculated as $\mathbf{BD}$.

2 Proof of Remark 1

Before proving Remark 1, we first give problem (15) in this article corresponding to the problem (1) here, problem (17) in this article corresponding to the problem (6) here, and Remark 1 as follows:

$$\begin{aligned} &\max_{\widehat{\mathbf{W}}} \mathbf{Tr}\left(\widehat{\mathbf{W}}^{\mathrm{T}}\left(\mathbf{S}_{\mathbf{E}}\widehat{\mathbf{W}}_0(\widehat{\mathbf{W}}_0^{\mathrm{T}}\mathbf{S}_{\mathbf{E}}\widehat{\mathbf{W}}_0)^{\dagger}\widehat{\mathbf{W}}_0^{\mathrm{T}}\mathbf{S}_{\mathbf{E}}\right)\widehat{\mathbf{W}}\right) \\ &\text{s.t. } \widehat{\mathbf{W}}^T\widehat{\mathbf{W}} = \mathbf{I}_{\mathrm{m}\times\mathrm{m}}, \|\widehat{\mathbf{W}}\|_{2,0} = \mathrm{k}. \end{aligned} \tag{6}$$

Remark 1. *Suppose that the objective functions of problem* (1) *and problem* (6) *are* $\mathcal{J}_R(\widehat{\mathbf{W}})$ *and* $\mathcal{J}_S(\widehat{\mathbf{W}})$ *respectively, at any point* $\widehat{\mathbf{W}}^T\widehat{\mathbf{W}} = \mathbf{I}_{\mathrm{m}\times\mathrm{m}}, \|\widehat{\mathbf{W}}\|_{2,0} = \mathrm{k}$*, we have* $\mathcal{J}_R(\widehat{\mathbf{W}}) \geqslant \mathcal{J}_S(\widehat{\mathbf{W}})$ *and* $\mathcal{J}_R(\widehat{\mathbf{W}}_0) = \mathcal{J}_S(\widehat{\mathbf{W}}_0)$.

The Proof of Remark 1. Let $\mathcal{J}_S(\widehat{\mathbf{W}}) = \mathrm{Tr}\left(\widehat{\mathbf{W}}^T\mathbf{S}_{\mathbf{E}}\widehat{\mathbf{W}}_0\left(\widehat{\mathbf{W}}_0^T\mathbf{S}_{\mathbf{E}}\widehat{\mathbf{W}}_0\right)^{\dagger}\widehat{\mathbf{W}}_0^T\mathbf{S}_{\mathbf{E}}\widehat{\mathbf{W}}\right) = \mathrm{Tr}\left(\mathbf{\Phi}\mathbf{\Psi}\right)$, where $\mathbf{\Phi} = \mathbf{S}_{\mathbf{E}}{}^{\frac{1}{2}}\widehat{\mathbf{W}}_0\left(\widehat{\mathbf{W}}_0^T\mathbf{S}_{\mathbf{E}}\widehat{\mathbf{W}}_0\right)^{\dagger}\widehat{\mathbf{W}}_0^T\mathbf{S}_{\mathbf{E}}{}^{\frac{1}{2}}$ and $\mathbf{\Psi} = \mathbf{S}_{\mathbf{E}}{}^{\frac{1}{2}}\widehat{\mathbf{W}}\widehat{\mathbf{W}}^T\mathbf{S}_{\mathbf{E}}{}^{\frac{1}{2}}$. According to **Theorem 4.3.53** and **Theorem 1.3.22** in the Book[2], we have

$$\mathrm{Tr}\left(\mathbf{\Phi}\mathbf{\Psi}\right) \leq \sum_{\mathrm{i}=1}^{\mathrm{d}} \lambda_{\mathrm{i}}\left(\mathbf{\Phi}\right)\lambda_{\mathrm{i}}\left(\mathbf{\Psi}\right) \leq \sum_{\mathrm{i}=1}^{\mathrm{m}} \lambda_{\mathrm{i}}\left(\mathbf{\Psi}\right). \tag{7}$$

Since $\mathrm{rank}\left(\mathbf{\Psi}\right) \leq \mathrm{rank}\left(\widehat{\mathbf{W}}\right) = m$, $\sum_{\mathrm{i}=1}^{\mathrm{m}} \lambda_{\mathrm{i}}\left(\mathbf{\Psi}\right) = \mathrm{Tr}\left(\mathbf{\Psi}\right) = \mathcal{J}_R(\widehat{\mathbf{W}})$, which complete the proof of $\mathcal{J}_R(\widehat{\mathbf{W}}) \geqslant \mathcal{J}_S(\widehat{\mathbf{W}})$. Moreover, $\mathcal{J}_R(\widehat{\mathbf{W}}_0) = \mathcal{J}_S(\widehat{\mathbf{W}}_0)$ holds since $\mathbf{A} = \mathbf{A}\mathbf{A}^{\dagger}\mathbf{A}$. This completes the proof of Reamrk 1. □

[2]Roger A Horn and Charles R Johnson. Matrix Analysis. Cambridge University Press, 1990.