Sparse sufficient dimension reduction

Qing Mai and Xin Zhang

March 6, 2018

Many dimension reduction methods finds the r-dimensional, column space of $\hat{\mathbf{M}}^{-1}\hat{\mathbf{U}}$, where $\hat{\mathbf{M}} \in \mathbb{R}^{p\times p}, \hat{\mathbf{U}} \in \mathbb{R}^{p\times K}$, and $r \leq K$. Define $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r)$ as a set of basis for this r-dimensional space. Note that $\mathbf{B}$ is not identifiable itself, but its column space is. In high dimensions, we can solve the following problem to estimate $\mathbf{B}$:

$$\hat{\mathbf{B}} = \arg\min_{\mathbf{B}\in\mathbb{R}^{p\times K}} \sum_{k=1}^{K}\{\boldsymbol{\beta}_k^{\mathrm{T}}\hat{\mathbf{M}}\boldsymbol{\beta}_k - 2\boldsymbol{\beta}_k^{\mathrm{T}}\hat{\mathbf{u}}_k\} + \lambda_1 \sum_j \sqrt{\sum_{k=1}^{K} b_{kj}^2} + \lambda_2 \|\mathbf{B}\|_* \tag{1}$$

where $\hat{\mathbf{u}}_k$ is the k'th column of $\hat{\mathbf{U}}$. The estimate $\hat{\mathbf{B}}$ is of a different dimension than $\mathbf{B}$, but its column space should be close to that of $\mathbf{B}$.

1 Algorithm

Because (1) is convex, we can use the ADMM algorithm to solve it. Consider the augmented problem:

$$(\hat{\mathbf{B}}, \hat{\mathbf{C}}) = \arg\min_{\mathbf{B}\in\mathbb{R}^{p\times K}, \mathbf{C}\in\mathbb{R}^{p\times K}, \gamma} \sum_{k=1}^{K}\{\boldsymbol{\beta}_k^{\mathrm{T}}\hat{\mathbf{M}}\boldsymbol{\beta}_k - 2\boldsymbol{\beta}_k^{\mathrm{T}}\hat{\mathbf{u}}_k\} + \lambda_1 \sum_j \sqrt{\sum_{k=1}^{K} b_{kj}^2} + \lambda_2 \|\mathbf{C}\|_* \tag{2}$$

$$\text{s.t. } \mathbf{B} = \mathbf{C} \tag{3}$$

Write the Lagrange:

$$L_\gamma(\mathbf{B}, \mathbf{C}, \boldsymbol{\mu}) = \sum_{k=1}^{K}\{\boldsymbol{\beta}_k^{\mathrm{T}}\hat{\mathbf{M}}\boldsymbol{\beta}_k - 2\boldsymbol{\beta}_k^{\mathrm{T}}\hat{\mathbf{u}}_k\} + \lambda_1 \sum_j \sqrt{\sum_{k=1}^{K} b_{ij}^2} + \lambda_2 \|\mathbf{C}\|_* + \langle \boldsymbol{\mu}, \mathbf{B} - \mathbf{C}\rangle + \frac{\gamma}{2}\|\mathbf{B} - \mathbf{C}\|_F^2 \tag{4}$$

where $\boldsymbol{\mu} \in \mathbb{R}^{p\times K}$ and $\gamma > 0$ is a small constant.

Then we have that

$$(\hat{\mathbf{B}}, \hat{\mathbf{C}}) = \arg \min_{\mathbf{B}\in\mathbb{R}^{p\times K}, \mathbf{C}\in\mathbb{R}^{p\times K}, \boldsymbol{\mu}} L_\gamma(\mathbf{B}, \mathbf{C}, \boldsymbol{\mu}) \tag{5}$$

We can solve (5) by iteratively solving the following problems:

$$\begin{aligned}
\mathbf{B}^{t+1} &= \arg\min_{\mathbf{B}} L_\gamma(\mathbf{B}, \mathbf{C}^t, \boldsymbol{\mu}^t) && (6)\\
\mathbf{C}^{t+1} &= \arg\min_{\mathbf{C}} L_\gamma(\mathbf{B}^{t+1}, \mathbf{C}, \boldsymbol{\mu}^t) && (7)\\
\boldsymbol{\mu}^{t+1} &= \boldsymbol{\mu}^t + \gamma(\mathbf{B}^{t+1} - \mathbf{C}^{t+1}) && (8)
\end{aligned}$$

Now we discuss how to solve (6) & (7). For (6), note that, if we fix $\mathbf{C}^t$, $\boldsymbol{\mu}^t$, we have

$$\begin{aligned}
& L_\gamma(\mathbf{B}, \mathbf{C}^t, \boldsymbol{\mu}^t) \\
= \; & \sum_{k=1}^{K} \{\boldsymbol{\beta}_k^{\mathrm{T}}(\hat{\mathbf{M}} + \gamma \mathbf{I})\boldsymbol{\beta}_k - 2\boldsymbol{\beta}_k^{\mathrm{T}}(\hat{\mathbf{u}}_k - \frac{1}{2}\boldsymbol{\mu}_k^t + \frac{\gamma}{2}\mathbf{c}_k^t)\} + \lambda_1 \sum_j \sqrt{\sum_{k=1}^{K} b_{ij}^2} + Const && (9)
\end{aligned}$$

where $\boldsymbol{\mu}_k^t$ is the kth column of $\boldsymbol{\mu}^t$ and $\mathbf{c}_k$ is the kth column of $\mathbf{C}^t$.

Therefore, (6) reduces to

$$\arg\min_{\mathbf{B}} \sum_{k=1}^{K} \{\boldsymbol{\beta}_k^{\mathrm{T}}(\hat{\mathbf{M}} + \gamma \mathbf{I})\boldsymbol{\beta}_k - 2\boldsymbol{\beta}_k^{\mathrm{T}}(\hat{\mathbf{u}}_k - \frac{1}{2}\boldsymbol{\mu}_k^t + \frac{\gamma}{2}\mathbf{c}_k^t)\} + \lambda_1 \sum_j \sqrt{\sum_{k=1}^{K} b_{ij}^2} \tag{10}$$

which can be solved by `msda`.

For (7),

$$\begin{aligned}
L_\gamma(\mathbf{B}^{t+1}, \mathbf{C}, \boldsymbol{\mu}^t) &= \lambda_2\|\mathbf{C}\|_* + \langle \boldsymbol{\mu}^t, \mathbf{B}^{t+1} - \mathbf{C}\rangle + \frac{\gamma}{2}\|\mathbf{B}^{t+1} - \mathbf{C}\|_F^2 + Const && (11)\\
&= \lambda_2\|\mathbf{C}\|_* + \langle \boldsymbol{\mu}^t, \mathbf{B}^{t+1} - \mathbf{C}\rangle + \frac{\gamma}{2}Tr((\mathbf{B}^{t+1} - \mathbf{C})^{\mathrm{T}}(\mathbf{B}^{t+1} - \mathbf{C})) + Const && (12)\\
&= \lambda_2\|\mathbf{C}\|_* + Tr(\frac{\gamma}{2}(\mathbf{B}^{t+1})^{\mathrm{T}}\mathbf{B}^{t+1} - \gamma(\mathbf{B}^{t+1})^{\mathrm{T}}\mathbf{C} + \frac{\gamma}{2}\mathbf{C}^{\mathrm{T}}\mathbf{C} + (\boldsymbol{\mu}^t)^{\mathrm{T}}\mathbf{B}^{t+1} - (\boldsymbol{\mu}^t)^{\mathrm{T}}\mathbf{C}) && (13)\\
&= \lambda_2\|\mathbf{C}\|_* + Tr(\frac{\gamma}{2}(\mathbf{C}^{\mathrm{T}}\mathbf{C} - 2(\mathbf{B}^{t+1} + \gamma^{-1}\boldsymbol{\mu}^t)^{\mathrm{T}}\mathbf{C})) + Const && (14)\\
&= \lambda_2\|\mathbf{C}\|_* + \frac{\gamma}{2}\|\mathbf{C} - (\mathbf{B}^{t+1} + \gamma^{-1}\boldsymbol{\mu}^t)\|_F^2 + Const && (15)
\end{aligned}$$

which is equivalent to soft-thresholding the singular values of $\mathbf{B}^{t+1} + \gamma^{-1}\boldsymbol{\mu}^t$ by $\frac{\lambda_2}{\gamma}$.

Conjecture: the one-step solution achieves both rank selection consistency and variable selection consistency.