Distributed Kernel Machine with Orthogonality

Byung-Jun Kim
Department of Statistics

1 How to distribute the matrix of covariate?

1.1 Blocking $\mathbb{X}$ over p

1.1.1 Polynomial Kernel

Recall the equation for the polynomial kernel denoted by K_p with a $n \times p$ matrix, $\boldsymbol{X}$ and a $p \times 1$ vector, $\boldsymbol{\xi}$ such that

$$K_p(\boldsymbol{X}, \boldsymbol{\xi}) = \mathbb{X}^T \big(\mathbf{1}\mathbf{1}^T \otimes \Omega\big)\mathbb{X} \in \mathbb{R}^{n \times n},$$

where

$$\Omega(\boldsymbol{\xi}) = \begin{pmatrix} \xi_1 & 0 & \cdots & 0 \\ 0 & \xi_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \xi_p \end{pmatrix} \in \mathbb{R}^{p \times p}.$$

and

$$\mathbb{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_{n-1} & \boldsymbol{x}_n \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & & \boldsymbol{x}_{n-1} & \boldsymbol{x}_n \\ \vdots & & \ddots & & \vdots \\ & & & & \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_{n-1} & \boldsymbol{x}_n \end{bmatrix} \in \mathbb{R}^{np \times n}.$$

Thus, the (l, m)th element of the matrix K_p can be expressed as

$$K_p(\boldsymbol{X})_{lm} = \sum_{j=1}^{p} \xi_j x_{lj} x_{mj} \equiv \boldsymbol{x}_l^T \Omega(\boldsymbol{\xi}) \boldsymbol{x}_m \in \mathbb{R}^1(\text{scalar}), \qquad 1 \le l, m \le n.$$

If we distribute the data matrix $\boldsymbol{X}$ among s blocks, then we have

$$\boldsymbol{X} = (\boldsymbol{X}^{(1)} | \boldsymbol{X}^{(2)} | \cdots | \boldsymbol{X}^{(s)})$$

where $X^{(i)}$s are $n \times p_i$ matrices with $p = \sum_{i=1}^{s} p_i$. Similarly, $\mathbb{X}$ can be divided by

$\mathbb{X} = (\mathbb{X}^{(1)} | \mathbb{X}^{(2)} | \cdots | \mathbb{X}^{(s)})$

$$\Rightarrow \begin{bmatrix} \boldsymbol{x}_1 = (\boldsymbol{x}_1^{(1)} | \boldsymbol{x}_1^{(2)} | \cdots | \boldsymbol{x}_1^{(s)}) & \boldsymbol{x}_2 = (\boldsymbol{x}_2^{(1)} | \boldsymbol{x}_2^{(2)} | \cdots | \boldsymbol{x}_2^{(s)}) & \cdots & \boldsymbol{x}_n = (\boldsymbol{x}_n^{(1)} | \boldsymbol{x}_n^{(2)} | \cdots | \boldsymbol{x}_n^{(s)}) \\ (\boldsymbol{x}_1^{(1)} | \boldsymbol{x}_1^{(2)} | \cdots | \boldsymbol{x}_1^{(s)}) & (\boldsymbol{x}_2^{(1)} | \boldsymbol{x}_2^{(2)} | \cdots | \boldsymbol{x}_2^{(s)}) & & (\boldsymbol{x}_n^{(1)} | \boldsymbol{x}_n^{(2)} | \cdots | \boldsymbol{x}_n^{(s)}) \\ \vdots & & \ddots & \vdots \\ & & & \\ (\boldsymbol{x}_1^{(1)} | \boldsymbol{x}_1^{(2)} | \cdots | \boldsymbol{x}_1^{(s)}) & (\boldsymbol{x}_2^{(1)} | \boldsymbol{x}_2^{(2)} | \cdots | \boldsymbol{x}_2^{(s)}) & \cdots & (\boldsymbol{x}_n^{(1)} | \boldsymbol{x}_n^{(2)} | \cdots | \boldsymbol{x}_n^{(s)}) \end{bmatrix} \in \mathbb{R}^{np \times n}.$$

Note that blocking of $\mathbb{X}$ is to partition each of the vectors inside the matrix into s sub-vectors. For the precision matrix, Ω, it is also distributed as

$$\Omega \Rightarrow \Omega^{(1)} = \begin{pmatrix} \xi_1 & 0 & \cdots & 0 \\ 0 & \xi_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \xi_{p_1} \end{pmatrix}, \Omega^{(2)} = \begin{pmatrix} \xi_{p_1+1} & 0 & \cdots & 0 \\ 0 & \xi_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \xi_{p_1+p_2} \end{pmatrix}, \cdots, \Omega^{(s)} = \begin{pmatrix} \xi_{\sum_{i=1}^{s-1} p_i+1} & 0 & \cdots & 0 \\ 0 & \xi_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \xi_{p=\sum_{i=1}^{s} p_i} \end{pmatrix},$$

where $\Omega^{(i)}$s are $p_i \times p_i$ matrices with $i = 1, ..., s$. After using the partition of the data matrix and the precision matrix, we can perform the distributed computation of the kernel matrix such that

$$\begin{aligned} K_p(\boldsymbol{X}, \boldsymbol{\xi}) &= \mathbb{X}^{(1)T}\big(\mathbf{1}\mathbf{1}^T \otimes \Omega^{(1)}\big)\mathbb{X}^{(1)} + \cdots + \mathbb{X}^{(s)T}\big(\mathbf{1}\mathbf{1}^T \otimes \Omega^{(s)}\big)\mathbb{X}^{(s)} \\ &= \sum_{i=1}^{s} \mathbb{X}^{(i)T}\big(\mathbf{1}\mathbf{1}^T \otimes \Omega^{(i)}\big)\mathbb{X}^{(i)} \in \mathbb{R}^{n\times n}. \end{aligned}$$

1.1.2 Guassian Kernel

Recall the equation for Gaussian kernel denotey by K_G based on a $n \times p$ matirx, $\boldsymbol{X}$ and a $p \times 1$ vector, $\boldsymbol{\xi}$ such that

$$K_G(\boldsymbol{\xi}, \boldsymbol{X}) = \exp\Big(- \mathbb{X}^{*T}\big(\mathbf{1}\mathbf{1}^t \otimes \Omega\big)\mathbb{X}^*\Big),$$

where

$$\mathbb{X}^* = \begin{bmatrix} \boldsymbol{x}^*_{1,1} & \boldsymbol{x}^*_{1,2} & \cdots & \boldsymbol{x}^*_{1,n-1} & \boldsymbol{x}^*_{1,n} \\ \boldsymbol{x}^*_{2,1} & \boldsymbol{x}^*_{2,2} & & \boldsymbol{x}^*_{2,n-1} & \boldsymbol{x}^*_{2,n} \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{x}^*_{n,1} & \boldsymbol{x}^*_{n,2} & \cdots & \boldsymbol{x}^*_{n,n-1} & \boldsymbol{x}^*_{n,n} \end{bmatrix} \in \mathbb{R}^{np\times n}.$$

and $\boldsymbol{X}^*$ consists of $\boldsymbol{x}^*_{l,m} = \big[(x_{1,l} - x_{1,m}), ..., (x_{p,l} - x_{p,m})\big]^T$ which is a $p \times 1$ vector. In the same way as the partition of the data matrix for the polynomial kernel, we can distribute the data matrix $\boldsymbol{X}^*$ as

$$\boldsymbol{X}^* = (\boldsymbol{X}^{*(1)}|\boldsymbol{X}^{*(2)}|\cdots|\boldsymbol{X}^{*(s)})$$

where $X^{(i)}$s are $n \times p_i$ matrices with $p = \sum_{i=1}^{s} p_i$. Similarly, $\mathbb{X}$ can be divided by

$$\mathbb{X}^* = (\mathbb{X}^{*(1)}|\mathbb{X}^{*(2)}|\cdots|\mathbb{X}^{*(s)})$$

$$\Rightarrow \begin{bmatrix} \boldsymbol{x}_1 = (\boldsymbol{x}_1^{*(1)}|\boldsymbol{x}_1^{*(2)}|\cdots|\boldsymbol{x}_1^{*(s)}) & \boldsymbol{x}_2 = (\boldsymbol{x}_2^{*(1)}|\boldsymbol{x}_2^{*(2)}|\cdots|\boldsymbol{x}_2^{*(s)}) & \cdots & \boldsymbol{x}_n = (\boldsymbol{x}_n^{*(1)}|\boldsymbol{x}_n^{*(2)}|\cdots|\boldsymbol{x}_n^{*(s)}) \\ (\boldsymbol{x}_1^{*(1)}|\boldsymbol{x}_1^{*(2)}|\cdots|\boldsymbol{x}_1^{*(s)}) & (\boldsymbol{x}_2^{*(1)}|\boldsymbol{x}_2^{*(2)}|\cdots|\boldsymbol{x}_2^{*(s)}) & & (\boldsymbol{x}_n^{*(1)}|\boldsymbol{x}_n^{*(2)}|\cdots|\boldsymbol{x}_n^{*(s)}) \\ \vdots & & \ddots & \vdots \\ (\boldsymbol{x}_1^{*(1)}|\boldsymbol{x}_1^{*(2)}|\cdots|\boldsymbol{x}_1^{*(s)}) & (\boldsymbol{x}_2^{*(1)}|\boldsymbol{x}_2^{*(2)}|\cdots|\boldsymbol{x}_2^{*(s)}) & \cdots & (\boldsymbol{x}_n^{*(1)}|\boldsymbol{x}_n^{*(2)}|\cdots|\boldsymbol{x}_n^{*(s)}) \end{bmatrix} \in \mathbb{R}^{np\times n}.$$

Then, the distributed computation for Guassian kernel is given by

$$\begin{aligned} K_G &= \exp\Big\{-\Big(\mathbb{X}^{*(1)T}\big(\mathbf{1}\mathbf{1}^T \otimes \Omega^{(1)}\big)\mathbb{X}^{*(1)} + \cdots + \mathbb{X}^{*(s)T}\big(\mathbf{1}\mathbf{1}^T \otimes \Omega^{(s)}\big)\mathbb{X}^{*(s)}\Big)\Big\} \\ &= \exp\Big(-\sum_{i=1}^{s} \mathbb{X}^{*(i)T}\big(\mathbf{1}\mathbf{1}^T \otimes \Omega^{(i)}\big)\mathbb{X}^{*(i)}\Big) \in \mathbb{R}^{n\times n}. \end{aligned}$$

1.2 Blocking $\mathbb{X}$ over n

The distributed kernel machine across n observations is a challenging problem because a lot of cross-products of two sub-matrices from the data matrix are needed. Further research is still needed to examine MPI and other distributed ways in detail.

2 Kernel Matrix with Conditional Dependence

To generalize effect of genes, it is necessary to take account of dependencies among them. Considering the conditional dependence among the components, the precision matrix can be expressed as

$$\Omega(\boldsymbol{\xi}) = \begin{pmatrix} \xi_{1,1} & \xi_{1,2} & \cdots & \xi_{1,p} \\ \xi_{2,1} & \xi_{2,2} & \xi_{2,3} & \vdots \\ \vdots & \xi_{3,2} & \ddots & \xi_{3,p} \\ \xi_{p,1} & \cdots & \xi_{p,p-1} & \xi_{p,p} \end{pmatrix} \in \mathbb{R}^{p\times p},$$

where the matrix might have nonzero off-diagonals and be symmetric such that $\xi_{jj'} = \xi_{j'j}$. In this case, the blocking of the data matrix over p is not applicable because of the dependence. The distributed kernel machine computing with dependency is another challenging problem.