Exploration of Covariance Matrix and Kernel Matrix

Byung-Jun Kim

June 4, 2019

1 Covariance Matrix and Kernel Matrix

1.1 Covariance Matrix

- Let $\boldsymbol{X}$ denote $n \times p$ matrix and $\boldsymbol{S}$ denote sample covariance matrix. Then, $\boldsymbol{S}$ is expressed as

$$n\boldsymbol{S} = \boldsymbol{X}^T(\boldsymbol{I} - n^{-1}\mathbf{1}\mathbf{1}^T)\boldsymbol{X} = \boldsymbol{X}^T\boldsymbol{X} - n^{-1}\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T \in \mathbb{R}^{p\times p},$$

  where $\bar{\boldsymbol{x}} = n^{-1}\mathbf{1}^T\boldsymbol{X} \in \mathbb{R}^p$

- For distributed computing of the covariance matrix, we split the matrix $\boldsymbol{X}$ into several blocks such that

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_s \end{bmatrix},$$

  where $\boldsymbol{X}_i$ are $n_i \times p$ matrices with $n = \sum_{i=1}^{s} n_i$. It is not necessary to make the blocks balanced with respect to the number of the observations.

- Using the idempotent centering operator denoted by $\boldsymbol{C}$, we can express the middle matrix of $\boldsymbol{S}$ as

$$
\begin{aligned}
\boldsymbol{C} &= [\boldsymbol{I} - n^{-1}\mathbf{1}\mathbf{1}^T] \\
&= \left[\boldsymbol{I} - \begin{pmatrix} n_1^{-1}\mathbf{1}\mathbf{1}^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & n_2^{-1}\mathbf{1}\mathbf{1}^T & & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & n_s^{-1}\mathbf{1}\mathbf{1}^T \end{pmatrix}\right] + \\
&\quad \left[\begin{pmatrix} n_1^{-1}\mathbf{1}\mathbf{1}^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & n_2^{-1}\mathbf{1}\mathbf{1}^T & & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & n_s^{-1}\mathbf{1}\mathbf{1}^T \end{pmatrix} - n^{-1}\mathbf{1}\mathbf{1}^T\right] \\
&= \boldsymbol{C}_w + \boldsymbol{C}_b,
\end{aligned}
$$

where $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{C}_w\boldsymbol{C}_b = \mathbf{0}$.

- Thus, we can rewrite the sample covariance matrix $\boldsymbol{S}$ as

$$
\begin{aligned}
n\boldsymbol{S} &= \boldsymbol{X}^T\boldsymbol{C}\boldsymbol{X} \\
&= \boldsymbol{X}^T\boldsymbol{C}_w\boldsymbol{X} + \boldsymbol{X}^T\boldsymbol{C}_b\boldsymbol{X} \\
&= \sum_{i=1}^{s} \boldsymbol{X}_i^T(\boldsymbol{I} - n_i^{-1}\mathbf{1}\mathbf{1}^T)\boldsymbol{X}_i + \boldsymbol{X}^T\boldsymbol{C}_b\boldsymbol{X} \\
&= \sum_{i=1}^{s} n_i\boldsymbol{S}_i + \boldsymbol{X}^T\boldsymbol{C}_b\boldsymbol{X}.
\end{aligned}
$$

1.2 Kernel Matrix in Orthogonal Setting

Following the notation for the covariance matrix, $\boldsymbol{X}$ denotes $n \times p$ matrix. K denotes $n \times n$ positive (semi-)definite kernel matrix $k(\boldsymbol{z}_l, \boldsymbol{z}_k)$, $1 \le l, k \le n$ as follows:

$$
K = \begin{bmatrix} k(\boldsymbol{x}_1,\boldsymbol{x}_1) & k(\boldsymbol{x}_1,\boldsymbol{x}_2) & \cdots & k(\boldsymbol{x}_1,\boldsymbol{x}_n) \\ k(\boldsymbol{x}_2,\boldsymbol{x}_1) & k(\boldsymbol{x}_2,\boldsymbol{x}_2) & & \vdots \\ \vdots & & \ddots & \vdots \\ k(\boldsymbol{x}_n,\boldsymbol{x}_1) & \cdots & \cdots & k(\boldsymbol{x}_n,\boldsymbol{x}_n) \end{bmatrix} \in \mathbb{R}^{n\times n},
$$

where

the (l,k)th entry of K is expressed as:

1. Linear polynomial kernel: $k(\boldsymbol{x}_l, \boldsymbol{x}_k) = \sum_{j=1}^{p} \xi_j x_{lj} x_{kj} = \sum_{j=1}^{p} \xi_j s_{lk}^{j}$,

2. Gaussian kernel: $k(\boldsymbol{x}_l, \boldsymbol{x}_k) = \exp\left(\sum_{j=1}^{p} -\xi_j (x_{lj} - x_{kj})^2\right) = \exp\left(\sum_{j=1}^{p} \xi_j s_{lk}^{j}\right)$,

where $\xi_j,\ j = 1, ..., p$ is a nonnegative garrote.

1.2.1 Linear Polynomial Kernel

- The (l,m)th element of the polynomial kernel matrix is

$$K(\boldsymbol{X})_{lk} = \sum_{j=1}^{p} \xi_j x_{lj} x_{kj} \equiv \boldsymbol{x}_l^T \Omega(\boldsymbol{\xi}) \boldsymbol{x}_k \in \mathbb{R}^1(\text{scalar}), \qquad 1 \leq l, k \leq n$$

  where $\Omega(\boldsymbol{\xi})$ is the $p \times p$ diagonal matrix with $\boldsymbol{\xi} = (\xi_1, ..., \xi_p)$ such as

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

- For the matrix representation, K can be expressed as follows:

$$K(\boldsymbol{\xi}^*, \boldsymbol{X}) = \mathbb{X}^T \left(\mathbf{1}\mathbf{1}^t \otimes \Omega\right) \mathbb{X},$$

  where

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

  and

$$\mathbf{1}\mathbf{1}^T \otimes \Omega = \begin{pmatrix} \Omega & \Omega & \cdots & \Omega \\ \Omega & \Omega & \cdots & \Omega \\ \vdots & & \ddots & \vdots \\ \Omega & \cdots & \Omega & \Omega \end{pmatrix} \in \mathbb{R}^{np \times np}.$$

  $\boldsymbol{x}_l$ is a $p \times 1$ vector with $l = 1, \ldots, n$.

1.2.2 Gaussian Kernel

- The (l,m)th element of the Gaussian kernel is

$$K(\boldsymbol{X})_{lk} = \exp\Big\{-\sum_{j=1}^{p}\xi_j(x_{lj}-x_{kj})^2\Big\} \equiv \exp\big\{-\boldsymbol{x}^{*T}_{l,k}\Omega^*(\boldsymbol{\xi})\boldsymbol{x}^*_{l,k}\big\}, \quad 1 \le l,k \le n$$

  where $\boldsymbol{x}^*_{l,k} = \big[(x_{l,1}-x_{k,1}),\ldots,(x_{l,p}-x_{k,p})\big]^T$ and $\Omega^*(\boldsymbol{\xi})$ is the $p \times p$ diagonal matrix for the distance vector, $\boldsymbol{x}^*_{l,k}$ with $\boldsymbol{\xi} = (\xi_1, ..., \xi_p)$.

- For the matrix representation, In this setting, we consider the polynomial kernel as

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

  where

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

$$\mathbf{1}\mathbf{1}^T \otimes \Omega = \begin{pmatrix} \Omega & \Omega & \cdots & \Omega \\ \Omega & \Omega & \cdots & \Omega \\ \vdots & & \ddots & \vdots \\ \Omega & \cdots & \Omega & \Omega \end{pmatrix} \in \mathbb{R}^{np\times np}.$$

  and

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

  where $\mathbb{X}$ is $np \times n$ consisting 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.

2 Similarity and Difference

2.1 Similarity

- Both covariance matrix and kernel matrix can be represented by cross-product.
- I need to think about how to decompose $\mathbb{I} \otimes \Omega$ similar to $\boldsymbol{C}$ of the covariance matrix.

2.2 Difference

- Dimensions of $\boldsymbol{S}$ and K are different; $\boldsymbol{S} \in \mathbb{R}^{p \times p}$ and $K \in \mathbb{R}^{n \times n}$.
- $\boldsymbol{C}$ and $\mathbb{I} \otimes \Omega$.