

Necessary condition for matrix-valued kernels

Miaoyan Wang, April 26, 2020

Theorem 0.1 (Necessary condition). *Suppose $\mathbf{K}: \mathbb{R}^{d' \times d} \times \mathbb{R}^{d' \times d} \mapsto \mathbb{R}^{d \times d}$ is a function that takes as input a pair of matrices and produces a matrix. Let $\{\mathbf{X}_i \in \mathbb{R}^{d' \times d}: i \in [n]\}$ denote a set of input matrices, and let \mathcal{K} denote an order-4 (n, n, d, d) -dimensional tensor,*

$$\mathcal{K} = \llbracket \mathcal{K}(i, i', p, p') \rrbracket, \quad \text{where } \mathcal{K}(i, i', p, p') \text{ is the } (p, p')\text{-th entry of the matrix } \mathbf{K}(\mathbf{X}_i, \mathbf{X}_{i'}).$$

Then, the factorization $\mathbf{K}(\mathbf{X}_i, \mathbf{X}_{i'}) = \mathbf{h}(\mathbf{X}_i)^T \mathbf{h}(\mathbf{X}_{i'})$ exists for some mapping \mathbf{h} , only if both of the following conditions hold:

- (1) *For every index $i \in [n]$, the matrix $\mathcal{K}(i, i, :, :) \in \mathbb{R}^{d \times d}$ is positive semidefinite.*
- (2) *For every index $p \in [d]$, the matrix $\mathcal{K}(:, :, p, p) \in \mathbb{R}^{n \times n}$ is positive semidefinite.*

Proof. (1) Let $i \in [n]$ be a fixed index. For any vector $\mathbf{a} \in \mathbb{R}^d$,

$$\mathbf{a}^T \mathcal{K}(i, i, :, :) \mathbf{a} = \mathbf{a}^T \mathbf{h}(\mathbf{X}_i)^T \mathbf{h}(\mathbf{X}_i) \mathbf{a} = \langle \mathbf{h}(\mathbf{X}_i) \mathbf{a}, \mathbf{h}(\mathbf{X}_i) \mathbf{a} \rangle = \|\mathbf{h}(\mathbf{X}_i) \mathbf{a}\|_2^2 \geq 0$$

- (2) Let $p \in [d]$ be a fixed index. We use $[\cdot]_{(k,p)}$ to denote the (k, p) -th entry of the matrix. For any vector $\mathbf{b} = (b_1, \dots, b_n)^T \in \mathbb{R}^n$,

$$\begin{aligned} \mathbf{b}^T \mathcal{K}(:, :, p, p) \mathbf{b} &= \sum_{ij} b_i b_j [\mathbf{h}(\mathbf{X}_i)^T \mathbf{h}(\mathbf{X}_j)]_{(p,p)} \\ &= \sum_{ij} b_i b_j \sum_k [\mathbf{h}(\mathbf{X}_i)]_{(k,p)} [\mathbf{h}(\mathbf{X}_j)]_{(k,p)} \\ &= \sum_k \left(\sum_i [\mathbf{h}(\mathbf{X}_i)]_{(k,p)} b_i \right) \left(\sum_j [\mathbf{h}(\mathbf{X}_j)]_{(k,p)} b_j \right) \\ &= \sum_k \left(\sum_i [\mathbf{h}(\mathbf{X}_i)]_{(k,p)} b_i \right)^2 \geq 0. \end{aligned}$$

□

Updated on April 29, 2020. Generalization of Mercer's theorem to matrix-valued predictors.

Definition 1 (Validity and Admissibility). We call the matrix-valued kernel \mathbf{K} a valid kernel if there exists a feature mapping \mathbf{h} such that $\mathbf{K}(\mathbf{X}, \mathbf{X}') = \mathbf{h}(\mathbf{X}) \mathbf{h}^T(\mathbf{X}')$ for all $\mathbf{X}, \mathbf{X}' \in \mathbb{R}^{d \times d'}$. We call \mathbf{K} an admissible kernel if the equality holds under the trace operation; i.e., $\text{tr} [\mathbf{K}(\mathbf{X}, \mathbf{X}')] = \text{tr} [\mathbf{h}(\mathbf{X}) \mathbf{h}^T(\mathbf{X}')] = \text{tr} [\mathbf{h}(\mathbf{X})^T \mathbf{h}(\mathbf{X}')] = \text{tr} [\mathbf{h}(\mathbf{X}')^T \mathbf{h}(\mathbf{X})]$ for all $\mathbf{X}, \mathbf{X}' \in \mathbb{R}^{d \times d'}$.

Theorem 0.2 (Characterization of admissible kernels). *Let $\mathbf{K}: \mathbb{R}^{d' \times d} \times \mathbb{R}^{d' \times d} \mapsto \mathbb{R}^{d \times d}$ be a function that takes as input a pair of matrices and produces a matrix. Define a function $\mathcal{F}: \mathbb{R}^{d' \times d} \times \mathbb{R}^{d' \times d} \mapsto \mathbb{R}$ as follows:*

$$\mathcal{F}(\mathbf{X}, \mathbf{X}') = \text{tr}[\mathbf{K}(\mathbf{X}, \mathbf{X}')], \text{ for all } \mathbf{X}, \mathbf{X}' \in \mathbb{R}^{d' \times d}.$$

Then, the following two statements are equivalent:

1. *The function \mathbf{K} is an admissible kernel.*
2. *The function \mathcal{F} is positive semidefinite.*

Remark 1. Recall that earlier we have defined two types of kernel \mathbf{K} :

- Hadamard-product type: $\mathbf{K}(\mathbf{X}, \mathbf{X}') = \underbrace{(\mathbf{X}^T \mathbf{X}' + \mathbf{1}\mathbf{1}^T) \circ \cdots \circ (\mathbf{X}^T \mathbf{X}' + \mathbf{1}\mathbf{1}^T)}_{d \text{ times}}.$
- Matrix-polynomial type: $\mathbf{K}(\mathbf{X}, \mathbf{X}') = (\mathbf{X}^T \mathbf{X}' + \mathbf{1}\mathbf{1}^T)^d.$

Theorem 0.2 provides a practical way to verify the non-existence of feature mapping for a given \mathbf{K} . Note that being admissible is a necessary condition for validity. Straightforward calculation shows that \mathcal{F} defined by the matrix-polynomial \mathbf{K} is not positive semidefinite, so the kernel \mathbf{K} is non-valid.

Remark 2. Can your algorithm go through with admissible kernels only? If so, then we have full characterization of the desired matrix-valued kernels..