Necessary condition for matrix-valued kernels

Miaoyan Wang, April 26, 2020

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

$$\mathcal{K}=[\![\mathcal{K}(i,i',p,p')]\!],\quad \textit{where } \mathcal{K}(i,i',p,p') \textit{ is the } (p,p')\textit{-th entry of the matrix } \boldsymbol{K}(\boldsymbol{X}_i,\boldsymbol{X}_{i'}).$$

Then, the factorization $\boldsymbol{K}(\boldsymbol{X}_i,\boldsymbol{X}_{i'})=\boldsymbol{h}(\boldsymbol{X}_i)^T\boldsymbol{h}(\boldsymbol{X}_{i'})$ exists for some mapping $\boldsymbol{h}$, 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 $\boldsymbol{a}\in\mathbb{R}^d$,

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

(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 $\boldsymbol{b}=(b_1,\ldots,b_n)^T\in\mathbb{R}^n$,

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

□

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

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

Theorem 0.2 (Characterization of admissible kernels)**.** *Let $\boldsymbol{K}: \mathbb{R}^{d' \times d} \times \mathbb{R}^{d' \times d} \mapsto \mathbb{R}^{d \times d}$ 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:*

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

Then, the following two statements are equivalent:

1. *The function $\boldsymbol{K}$ is an admissible kernel.*

2. *The function $\mathcal{F}$ is positive semidefinite.*

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

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

- Matrix-polynomial type: $\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}') = (\boldsymbol{X}^T \boldsymbol{X}' + \mathbb{1}\mathbb{1}^T)^d$.

Theorems 0.1 and 0.2 can be used to disprove the existence of feature mapping for a given $\boldsymbol{K}$. Note that being admissiable is a necessary condition for validity. Straightforward calculation shows the non positive definiteness of $\mathcal{F}$ and therefore the non-validity of $\boldsymbol{K}$.

Remark 2. It seems that your algorithm still goes through with admissible kernels (verify?) If so, then we have full characterization of desired kernels for computations. The theoretical analysis may require valid kernels though...