

An equivalent formulation of matrix kernels (II)

Miaoyan Wang, Aug 3, 2020

1 Corrections to your section 1

Let $K(\cdot, \cdot)$ denote a usual kernel defined over vector pairs in \mathbb{R}^d . We use the shorthand $K(i, j) \stackrel{\text{def}}{=} K(\mathbf{X}_{i:}, \mathbf{X}'_{j:})$ to denote the kernel value evaluated on the vector pair $(\mathbf{X}_{i:}, \mathbf{X}'_{j:})$.

(Note: my projection matrix \mathbf{P} is the transpose of your \mathbf{P} .)

Proposition 1 (Rank-1 weights in Kernel). Define a kernel over matrix pairs, $\mathcal{K}(\mathbf{X}, \mathbf{X}') \stackrel{\text{def}}{=} \langle \mathbf{P}^T \Phi(\mathbf{X}) \mathbf{P}, \mathbf{P}^T \Phi(\mathbf{X}') \mathbf{P} \rangle$ for some rank-1 projection matrix $\mathbf{P} \in \mathbb{R}^{2 \times 1}$. Then, \mathcal{K} has an equivalent representation,

$$\mathcal{K}(\mathbf{X}, \mathbf{X}') = 2C \sum_{i,j} w_{ij} K(i, j), \quad (1)$$

where $\mathbf{W} = \mathbf{P} \mathbf{P}^T = \llbracket w_{ij} \rrbracket$ is a rank-1 weight matrix, and $C > 0$ is a normalizing constant.

Proof. By definition,

$$\begin{aligned} \mathcal{K}(\mathbf{X}, \mathbf{X}') &= \langle \mathbf{P}^T \Phi(\mathbf{X}) \mathbf{P}, \mathbf{P}^T \Phi(\mathbf{X}') \mathbf{P} \rangle \\ &= \langle \underbrace{\mathbf{P} \mathbf{P}^T}_{=: \mathbf{W}}, \Phi^T(\mathbf{X}) \underbrace{\mathbf{P} \mathbf{P}^T}_{=: \mathbf{W}} \Phi(\mathbf{X}') \rangle \end{aligned} \quad (2)$$

Both \mathbf{W} and $\Phi^T(\mathbf{X}) \mathbf{W} \Phi(\mathbf{X}')$ are d -by- d matrices. The (i, j) -th entry of $\Phi^T(\mathbf{X}) \mathbf{W} \Phi(\mathbf{X}')$ is

$$\begin{aligned} [\Phi^T(\mathbf{X}) \mathbf{W} \Phi(\mathbf{X}')]_{ij} &= \sum_{s,s'} [\Phi^T(\mathbf{X})]_{is} [\mathbf{W}]_{ss'} [\Phi(\mathbf{X}')]_{s'j} \\ &= \sum_{s,s'} w_{ss'} \langle (\phi(\mathbf{X}_{s:}), \phi(\mathbf{X}_{i:})), (\phi(\mathbf{X}'_{s':}), \phi(\mathbf{X}'_{j:})) \rangle \\ &= \sum_{s,s'} w_{ss'} (K(s, s') + K(i, j)) \\ &= CK(i, j) + \sum_{s,s'} w_{ss'} K(s, s'), \end{aligned} \quad (3)$$

where we have denoted the constant $C = \sum_{s,s'} w_{ss'} > 0$. Plugging (3) into (2) gives

$$\begin{aligned} \mathcal{K}(\mathbf{X}, \mathbf{X}') &= \sum_{i,j} w_{ij} [\Phi^T(\mathbf{X}) \mathbf{W} \Phi(\mathbf{X}')]_{ij} \\ &= C \sum_{i,j} w_{ij} K(i, j) + \left(\sum_{i,j} w_{ij} \right) \left(\sum_{s,s'} w_{ss'} K(s, s') \right) \\ &= 2C \sum_{i,j} w_{ij} K(i, j). \end{aligned}$$

□

Proposition 2 (Compatibility with row-wise-only mapping). Based on your Section 2, the row-wise-only mapping induces the following kernel,

$$\langle \Phi(\mathbf{X})\mathbf{P}, \Phi(\mathbf{X}')\mathbf{P} \rangle = \sum_{i,j} w_{ij} K(i,j), \text{ where } \mathbf{W} = \llbracket w_{ij} \rrbracket = \mathbf{P}^T \mathbf{P} \text{ is a low-rank p.s.d. matrix.}$$

This kernel is proportional to that in (1).

2 Commentary to your section 3

Proposition 3 (Isomorphic Mappings; From Mapping to Kernel). The following two mappings are isomorphic, in the sense that they induce the same kernel \mathcal{K} over matrix pairs.

- Mapping 1

$$\begin{aligned} \Phi_1 : \mathbb{R}^{d_1 \times d_2} &\rightarrow \mathcal{H}_r^{d_1} \times \mathcal{H}_c^{d_2} \\ \mathbf{X} &\mapsto (\Phi_r(\mathbf{X}), \Phi_c(\mathbf{X})) \stackrel{\text{def}}{=} (\phi_r(\mathbf{X}_{1:}), \dots, \phi_r(\mathbf{X}_{d_1:}), \phi_c(\mathbf{X}_{:1}), \dots, \phi_c(\mathbf{X}_{:d_2})) \end{aligned}$$

- Mapping 2

$$\begin{aligned} \Phi_2 : \mathbb{R}^{d_1 \times d_2} &\rightarrow (\mathcal{H}_r \times \mathcal{H}_c)^{d_1 \times d_2} \\ \mathbf{X} &\mapsto [\Phi_2(\mathbf{X})_{ij}], \quad \text{where } \Phi_2(\mathbf{X})_{ij} \stackrel{\text{def}}{=} (\phi_c(\mathbf{X}_{i:}), \phi_r(\mathbf{X}_{:j})) \end{aligned}$$

Proof. Using the similar argument in Proposition 1, we show that the kernel induced by (mapping 2 + low-rank coefficients) is

$$\begin{aligned} \mathcal{K} : \mathbb{R}^{d_1 \times d_2} \times \mathbb{R}^{d_1 \times d_2} &\rightarrow \mathbb{R} \\ \mathcal{K}(\mathbf{X}, \mathbf{X}') &\mapsto \sum_{i,j \in [d_1]} w_{ij}^{\text{row}} K_r(i,j) + \sum_{i,j \in [d_2]} w_{ij}^{\text{col}} K_c(i,j), \end{aligned} \quad (4)$$

where $\mathbf{W}^{\text{row}} = \llbracket w_{ij}^{\text{row}} \rrbracket = \frac{1}{c_1} \mathbf{P}_r \mathbf{P}_r^T$, $\mathbf{W}^{\text{col}} = \llbracket w_{ij}^{\text{col}} \rrbracket = \frac{1}{c_2} \mathbf{P}_c \mathbf{P}_c^T$ are some low-rank p.s.d. matrices, and $c_1 = \|\mathbf{1}_{d_1}^T \mathbf{P}_r\|_2^2 > 0$, $c_2 = \|\mathbf{1}_{d_2}^T \mathbf{P}_c\|_2^2 > 0$ are two normalizing constants.

Now, we consider the kernel induced by (mapping 1 + low-rank coefficients),

$$\begin{aligned} \mathcal{K}(\mathbf{X}, \mathbf{X}') &= \langle \Phi_r(\mathbf{X})\mathbf{P}_r, \Phi_r(\mathbf{X}')\mathbf{P}_r \rangle + \langle \Phi_c(\mathbf{X})\mathbf{P}_c, \Phi_c(\mathbf{X}')\mathbf{P}_c \rangle \\ &= \sum_{i,j \in [d_1]} w_{ij}^{\text{row}} K_r(i,j) + \sum_{i,j \in [d_2]} w_{ij}^{\text{col}} K_c(i,j), \end{aligned} \quad (5)$$

where $\mathbf{W}^{\text{row}} = \llbracket w_{ij}^{\text{row}} \rrbracket$, $\mathbf{W}^{\text{col}} = \llbracket w_{ij}^{\text{col}} \rrbracket$ are some low-rank p.s.d. matrices. □

Two important properties in the induced kernels (4) and (5):

1. [Additivity] The new kernel is a linear combination of regular row and column kernels;
2. [Low-rank p.s.d.] The weight matrices \mathbf{W}^{row} , \mathbf{W}^{col} are low-rank + p.s.d.

Conjecture 1 (From Kernel to Mapping). Let $\mathcal{K}(\cdot, \cdot)$ be a function that maps a pair of matrices to a real-value. Suppose $\mathcal{K}(\cdot, \cdot)$ satisfies the above two properties. Then, the kernel \mathcal{K} induces a decomposable feature mapping in that $\Phi(\mathbf{X}) = \Phi_r(\mathbf{X}) + \Phi_c(\mathbf{X})$, where, informally speaking, $\Phi_r(\cdot)$, $\Phi_c(\cdot)$ are the row- and column-wise mappings, respectively.

The decomposable mapping means the effects from rows and columns are additive/separable. Similar to an ANOVA model $Y_{ij} = \mu_i + \mu_j$ with marginal effects only. Additivity is a common assumption for matrix-based network analysis; see [?].

References

- [1] Peter D Hoff. Additive and multiplicative effects network models. *To appear in Statistical Science, arXiv preprint arXiv:1807.08038*, 2018.