CS-233 Theoretical Exercise 7

February 2024

1 Valid Kernels

Let us assume that a kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ is valid if there exists a mapping $\phi(\cdot)$ to the real domain such that $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ for all $\mathbf{x}_i$ and $\mathbf{x}_j$.

Question 1: Decide whether the following are valid kernels and explain why.

1. $k(\mathbf{x}_i, \mathbf{x}_j) = 4$
2. $k(\mathbf{x}_i, \mathbf{x}_j) = -4$
3. $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)^2$
4. $k(\mathbf{x}_i, \mathbf{x}_j) = \cos(\mathbf{x}_i - \mathbf{x}_j)$
5. $k(\mathbf{x}_i, \mathbf{x}_j) = \cos(\mathbf{x}_i)\sin(\mathbf{x}_j)$

Solution:

1. Yes. Let $\phi(\mathbf{x}) = 2$.
2. No. $k(\mathbf{x}, \mathbf{x})$ should not be negative.
3. Yes. It is a special case of polynomial kernel on page 22 of the slide.
4. Yes. Since $\cos(\mathbf{x}_i - \mathbf{x}_j) = \cos(\mathbf{x}_i)\cos(\mathbf{x}_j) + \sin(\mathbf{x}_i)\sin(\mathbf{x}_j)$, let $\phi(\mathbf{x}) = [\cos(\mathbf{x}), \sin(\mathbf{x})]$.
5. No. $k(\mathbf{x}, \mathbf{x})$ should not be negative.

2 Constructing Kernels

One powerful technique for constructing new kernels is to build them out of simpler kernels as building blocks. Assume that kernels $k_1(\mathbf{x}_i, \mathbf{x}_j)$ and $k_2(\mathbf{x}_i, \mathbf{x}_j)$ are valid.

Question 2: Show that the following kernels are valid.

1. $k(\mathbf{x}_i, \mathbf{x}_j) = ck_1(\mathbf{x}_i, \mathbf{x}_j)$, where $c > 0$ is a constant
2. $k(\mathbf{x}_i, \mathbf{x}_j) = k_1(\mathbf{x}_i, \mathbf{x}_j) + k_2(\mathbf{x}_i, \mathbf{x}_j)$
3. $k(\mathbf{x}_i, \mathbf{x}_j) = k_1(\mathbf{x}_i, \mathbf{x}_j)k_2(\mathbf{x}_i, \mathbf{x}_j)$

Solution:

1. Suppose $k_1(\mathbf{x}_i, \mathbf{x}_j) = \phi_1(\mathbf{x}_i)^T \phi_1(\mathbf{x}_j)$. Let $\phi(x) = \sqrt{c}\phi_1(\mathbf{x})$, then $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$.
2. $\phi(x) = [\phi_1(\mathbf{x}), \phi_2(\mathbf{x})]$, where $[\cdot, \cdot]$ means concatenation.

3. $k(\mathbf{x}_i, \mathbf{x}_j) = k_1(\mathbf{x}_i, \mathbf{x}_j)k_2(\mathbf{x}_i, \mathbf{x}_j) = \phi_1(\mathbf{x}_i)^T\phi_1(\mathbf{x}_j)\phi_2(\mathbf{x}_i)^T\phi_2(\mathbf{x}_j) = \sum_{m=1}^{M}\phi_{1,m}(\mathbf{x}_i)\phi_{1,m}(\mathbf{x}_j)\sum_{n=1}^{N}\phi_{2,n}(\mathbf{x}_i)\phi_{2,n}(\mathbf{x}_j) = \sum_{m=1}^{M}\sum_{n=1}^{N}\phi_{1,m}(\mathbf{x}_i)\phi_{1,m}(\mathbf{x}_j)\phi_{2,n}(\mathbf{x}_i)\phi_{2,n}(\mathbf{x}_j) = \sum_{m=1}^{M}\sum_{n=1}^{N}(\phi_{1,m}(\mathbf{x}_i)\phi_{2,n}(\mathbf{x}_i))(\phi_{1,m}(\mathbf{x}_j)\phi_{2,n}(\mathbf{x}_j))$.
It is a summation of valid kernels, so according to sub-question (2), it is also a valid kernel.

3 Kernelized Nearest Neighbor

The standard version of the Nearest Neighbor algorithm exploits the Euclidean distance $d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|$. If we have a valid kernel $k(\mathbf{x}_i, \mathbf{x}_j)$, then we can define a new distance based on the kernel as $d_k(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{k(\mathbf{x}_i - \mathbf{x}_j, \mathbf{x}_i - \mathbf{x}_j)}$. Suppose that we have three points on a plane $(0,1), (1,0), (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$.

Question 3: Find the nearest neighbor of $(0,0)$ under the following distances based on different kernels.

1. $k(\mathbf{x}_i, \mathbf{x}_j) = x_i^T x_j$
2. $k(\mathbf{x}_i, \mathbf{x}_j) = x_i^T \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} x_j$

Solution:

1. The distances to the three points are $1, 1, 1$. So all three points are the nearest neighbors.
2. The distances to the three points are $2, 1, \sqrt{2.5}$. So $(1,0)$ is the nearest neighbor.