CSE 151: Introduction to Machine Learning Spring 2018

Problem Set 5

Instructor: Kamalika Chaudhuri **Due on:**

Problem 1: 8 points

In the following problems, suppose that K, K_1 and K_2 are kernels with feature maps ϕ, ϕ^1 and ϕ^2. For the following functions $K'(x, z)$, state if they are kernels or not. If they are kernels, write down the corresponding feature map, in terms of ϕ, ϕ^1, ϕ^2 and c, c_1, c_2. If they are not kernels, prove that they are not.

1. $K'(x, z) = cK(x, z)$, for $c > 0$.
2. $K'(x, z) = cK(x, z)$, where $c < 0$, and there exists some x for which $K(x, x) > 0$.
3. $K'(x, z) = c_1 K_1(x, z) + c_2 K_2(x, z)$ for $c_1, c_2 > 0$.
4. $K'(x, z) = K_1(x, z)K_2(x, z)$.

Solution

1. Suppose $K(x, z) = \langle \phi(x), \phi(z) \rangle$ for some feature map ϕ, and let $\phi'(x) = \sqrt{c}\phi(x)$. Then, for all x and z,

   $$K'(x, z) = cK(x, z) = c\langle \phi(x), \phi(z) \rangle = \langle \sqrt{c}\phi(x), \sqrt{c}\phi(z) \rangle$$

   Therefore $K'(x, z)$ is a kernel corresponding to the feature map ϕ'.

2. Suppose x_0 is the x for which $K(x, x) > 0$. Consider the 1×1 kernel matrix $K' = K'(x_0, x_0)$ for the kernel K' and the data point x_0. Then, $K' = cK(x_0, x_0)$. If $z = 1$, then $z^\top K' z = cK(x_0, x_0) < 0$, which violates the kernel Positive Semi Definiteness (PSD) property. Thus K' is not a kernel.

3. Suppose $K_1(x, z) = \langle \phi^1(x), \phi^1(z) \rangle$ and $K_2(x, z) = \langle \phi^2(x), \phi^2(z) \rangle$. Then, for all x and z,

   $$\begin{aligned} K'(x, z) &= c_1\langle \phi^1(x), \phi^1(z) \rangle + c_2\langle \phi^2(x), \phi^2(z) \rangle = \langle \sqrt{c_1}\phi^1(x), \sqrt{c_1}\phi^1(z) \rangle + \langle \sqrt{c_2}\phi^2(x), \sqrt{c_2}\phi^2(z) \rangle \\ &= \langle \phi'(x), \phi'(z) \rangle \end{aligned}$$

   where $\phi'(x)$ is a concatenation of the feature maps $\sqrt{c_1}\phi^1(x)$ and $\sqrt{c_2}\phi^2(x)$. In other words, if the feature maps ϕ^1 and ϕ^2 have m_1 and m_2 coordinates respectively, then ϕ' has $m_1 + m_2$ coordinates; for any x, the first m_1 coordinates of $\phi'(x)$ are $\sqrt{c_1}\phi^1_1(x), \sqrt{c_1}\phi^1_2(x), \ldots, \sqrt{c_1}\phi^1_{m_1}(x)$ and the remaining m_2 coordinates of $\phi'(x)$ are $\sqrt{c_2}\phi^2_1(x), \sqrt{c_2}\phi^2_2(x), \ldots, \sqrt{c_2}\phi^2_{m_2}(x)$. Therefore $K'(x, z)$ is a kernel corresponding to the feature map ϕ'.

4. Suppose $K_1(x, z) = \langle \phi^1(x), \phi^1(z) \rangle$ and $K_2(x, z) = \langle \phi^2(x), \phi^2(z) \rangle$. If x and z are d-dimensional vectors, then, for all x and z,

   $$\begin{aligned} K'(x, z) &= K_1(x, z)K_2(x, z) = \langle \phi^1(x), \phi^1(z) \rangle \cdot \langle \phi^2(x), \phi^2(z) \rangle \\ &= \left(\sum_i \phi^1_i(x)\phi^1_i(z) \right) \cdot \left(\sum_j \phi^2_j(x)\phi^2_j(z) \right) = \sum_{i,j=1}^{d} (\phi^1_i(x)\phi^2_j(x)) \cdot (\phi^1_i(z)\phi^2_j(z)) \\ &= \langle \phi'(x), \phi'(z) \rangle \end{aligned}$$

where

$$\phi'(x) = \begin{bmatrix} \phi_1^1(x)\phi_1^2(x) \\ \phi_1^1(x)\phi_2^2(x) \\ \phi_2^1(x)\phi_1^2(x) \\ \phi_1^1(x)\phi_3^2(x) \\ \phi_2^1(x)\phi_2^2(x) \\ \phi_3^1(x)\phi_1^2(x) \\ \vdots \end{bmatrix} \tag{1}$$

That is, ϕ' is a $d^2 \times 1$ feature map, which has a coordinate $\phi_{(i,j)}(\cdot)$ corresponding to each pair (i,j), $1 \le i,j \le d$, where $\phi_{(i,j)}(x) = \phi_i^1(x)\phi_j^2(x)$. Thus $K'(x,z)$ is a kernel corresponding to the feature map ϕ'.

Problem 2: 14 points

For the following functions $K(x,z)$, state if it is a kernel or not. If the function is a kernel, then write down its feature map. If it is not a kernel, prove that it is not one.

1. $x = [x_1, x_2]$, $z = [z_1, z_2]$, x_1, x_2, z_1, z_2 are real numbers. $K(x,z) = x_1 z_2$.
2. Let $x = [x_1, \dots, x_d]$, $z = [z_1, \dots, z_d]$, x_is and z_is are real numbers. $K(x,z) = 1 - \langle x, z\rangle$.
3. $x = [x_1, \dots, x_d]$, $z = [z_1, \dots, z_d]$, and f is a function. $K(x,z) = f(x_1,x_2)f(z_1,z_2)$.
4. $x = [x_1, \dots, x_d]$, $z = [z_1, \dots, z_d]$, x_is and z_is are integers between 0 and 100. $K(x,z) = \sum_{i=1}^d \min(x_i, z_i)$.
5. $x = [x_1, \dots, x_d]$, $z = [z_1, \dots, z_d]$, x_is and z_is are real numbers.

$$K(x,z) = (1 + x_1 z_1)(1 + x_2 z_2)\dots(1 + x_d z_d)$$

6. $x = [x_1, \dots, x_d]$, $z = [z_1, \dots, z_d]$, x_is and z_is are integers between 0 and 100. $K(x,z) = \sum_{i=1}^d \max(x_i, z_i)$.
7. x are z are documents with words from some dictionary D. $K(x,z)$ is the number of words that occur in both x and z, where each unique common word is counted once.

Solution

1. $K(x,z)$ is not a kernel.
   For $x = [1, -1]$, we have $K(x,x) = 1 \times -1 = -1$. The corresponding kernel matrix $K = -1$. For $v = 1$, $v^\top K v = -1 < 0$, which violates the PSD property. Thus K is not a kernel.

2. $K(x,z)$ is not a kernel.
   For $x = [2, 2, \cdots]$, we have $K(x,x) = 1 - \langle x, x\rangle = 1 - 4d$. The corresponding kernel matrix $K = 1 - 4d$. For $v = 1$, $v^\top K v = 1 - 4d < 0$, which violates the kernel PSD property for $d > 0$. Thus K is not a kernel.

3. $K(x,z)$ is a kernel corresponding to the feature map $\phi(x) = f(x_1, x_2)$.

4. $K(x,z)$ is a kernel.
   Let $K_i(x,z) = \min(x_i, z_i)$. From Problem 1, we know that the sum of two kernels K_1 and K_2 is also a kernel whose corresponding feature map is the concatenation of the feature maps corresponding to K_1 and K_2. Thus if we can find the feature maps for all $K_i(x,z)$, then we can get the feature map for $K(x,z)$ by concatenating these maps. Consider following feature map:

$$\phi_i(x) = [f_1(x_i), f_2(x_i), \dots, f_{100}(x_i)]^\top \tag{2}$$

where $f_k(t) = I(t \geq k) = \begin{cases} 1 & t \geq k \\ 0 & t < k \end{cases}$. Without loss of generality, suppose that $x_i \leq z_i$. Then $\phi_i(x) = [1, \ldots, 1, 0, \ldots, 0]^\top$ where only the first x_i entries are 1. Analogously, $\phi_i(z) = [1, \ldots, 1, 0, \ldots, 0]^\top$ where only the first z_i entries are 1. Then

$$\langle \phi_i(x), \phi_i(z) \rangle = \sum_{i=1}^{x_i} 1 \cdot 1 + \sum_{i=x_i+1}^{z_i} 0 \cdot 1 + \sum_{i=z_i+1}^{100} 0 \cdot 0 = x_i = \min(x_i, z_i)$$

Therefore $K_i(x, z)$ is a kernel corresponding to the feature map $\phi_i(x) = [f_1(x_i), f_2(x_i), \ldots, f_{100}(x_i)]^\top$, and $K(x, z)$ is a kernel corresponding to the feature map $\phi(x)$ which is a concatenation of the feature maps $\phi_1(x), \phi_2(x), \ldots, \phi_d(x)$.

5. $K(x, z)$ is a kernel.

   Let $K_i(x, z) = 1 + x_i z_i$, then $K(x, z) = \prod_{i=0}^{d} K_i(x)$. From Problem 1, we know that the product of two kernels is also a kernel. Since $K_i(x, z)$ is a kernel corresponding to the feature map $\phi_i(x) = [1, x_i]^\top$, $K(x, z)$ is also a kernel. More specifically, $K(x, z)$ is a kernel corresponding to the feature map $\phi(x)$, where for any x, $\phi(x)$ has 2^d coordinates, one corresponding to each subset S of $\{1, 2, \ldots, d\}$. $\phi_S(x)$, the coordinate of $\phi(x)$ corresponding to the set S is $\prod_{i \in S} x_i$. This kernel is called the *All Subsets* kernel.

6. $K(x, z)$ is not a kernel.
   One way to prove this is by showing a violation of the PSD property. Let $x = [0, \ldots, 0], z = [1, 0, \ldots, 0]$ and $v = [1, -1]^\top$. Then the kernel matrix

$$K = \begin{bmatrix} K(x, x) & K(x, z) \\ K(z, x) & K(z, z) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

   Thus, $v^\top A v = -1 < 0$, which violates positivity.

   Another nice way is through a violation of the Cauchy-Schwartz inequality. Consider $x = [0, \ldots, 0]$ and $z = [1, 0, \ldots, 0]$. Then $K(x, x) = 0$, $K(x, z) = K(z, z) = 1$, which violates Cauchy-Schwarz inequality – that is $K(x, z)^2 \geq K(x, x) \cdot K(z, z)$.

7. K is a kernel. The feature map ϕ has a coordinate for each word u in the dictionary D. Given a document x, the coordinate of $\phi(x)$ corresponding to word u, $\phi_u(x)$, is 1 if x contains the word u and 0 otherwise. Notice that this kernel is very similar to the string kernel we discussed in class.