Kernel Methods in Machine Learning
Homework 1

Master Data Science – Université Paris-Saclay

Peter MARTIGNY : peter.martigny@gmail.com

Wednesday, January 25th 2017

1 Kernel Examples

Are the following kernels positive definite ?

1. 1.

$$\forall x, y \in \mathbb{R}, \mathrm{K}_1(x, y) = 10^{xy},\ K_2(x, y) = 10^{x+y}$$

2. 2.

$$\forall x, y \in [0,\ 1), \mathrm{K}_3(x, y) = -log(1 - xy)$$

3. 3. Let X be a set and $f, g : X \to \mathbb{R}_+$ two non-negative functions :

$$\forall x, y \in X, \mathrm{K}_4(x, y) = min(f(x)g(y), f(y)g(x))$$

1. 1.
   We know that the function $(x, y) \mapsto xy$ is a positive definite kernel. Moreover, we know that a p.d. kernel multiplied by a positive constant is still a p.d. kernel. Hence, $(x, y) \mapsto xyln(10)$. is a p.d. kernel. We also know that the multiplication and the addition of p.d. kernels gives still a p.d. kernel, hence the exponential of $(x, y) \mapsto xyln(10)$ (which can be written as a Taylor infinite expansion, with strictly positive factors for each power of $(x, y) \mapsto xyln(10)$) is still a p.d. kernel, which means that K_1 is a p.d. kernel.

   Let $\Phi : x \mapsto 10^x$. Then, we have : $K_3(x, y) = \Phi(x)\Phi(y)$. According to the Aronszajin Theorem, this shows that K_3 is a p.d. kernel.
2. 2. Let us write the Taylor Expansion of $x \mapsto log(1 - x)$. We have :

$$-log(1 - x) = \sum_{i=1}^{\infty} \frac{x^k}{k}$$

   Hence, as $(x, y) \mapsto xy$ ias p.d. kernel, we know that $(x, y) \mapsto \sum_{i=1}^{\infty} \frac{(xy)^k}{k}$ is also a p.d. kernel. Therefore, $(x, y) \mapsto -log(1 - xy)$ is a positive definite kernel.
3. 3.
   Because f and g are non negative, we show with Aronszajin Theorem that the function $(x, y) \mapsto f(x)g(y)$ is a p.d. kernel. Let us show now that K_4 is also a p.d. kernel.
   Let $N \in \mathbb{N}$, $(\alpha_i, ..., \alpha_N) \in \mathbb{R}^N$ and $(x_i, ..., x_N) \in \mathbb{X}^N$. We have :

$$\sum_{i,j} \alpha_i \alpha_j min(f(x_i)g(x_j), f(y)g(x)) = \sum_{i,j} \alpha_i \alpha_j \delta_{f(x_i)g(x_j) < f(x_j)g(x_i)} f(x_i)g(x_j) + \sum_{i,j} \alpha_i \alpha_j \delta_{f(x_j)g(x_i) < f(x_i)g(x_j)} f(x_j)g(x_i)$$

   Hence, since we know that $(x, y) \mapsto f(x)g(y)$ is a p.d. kernel (Aronszajin theorem), we have that :

$$\sum_{i,j} \alpha_i \alpha_j \delta_{f(x_i)g(x_j)<f(x_j)g(x_i)} f(x_i)g(x_j) \geq 0$$

$$\sum_{i,j} \alpha_i \alpha_j \delta_{f(x_j)g(x_i)<f(x_i)g(x_j)} f(x_j)g(x_i) \geq 0$$

Hence we have :

$$\sum_{i,j} \alpha_i \alpha_j min(f(x_i)g(x_j), f(y)g(x)) \geq 0$$

Therefore, K_4 is a p.d. kernel.

2 Combining Kernels

1. For $x, y \in \mathbb{R}$, let

$$K_1(x, y) = (xy + 1)^2$$
$$K_2(x, y) = (xy - 1)^2$$

What is the RKHS of K_1 ? Of K_2 ? Of $K_1 + K_2$?

2. Let K_1 and K_2 be two positive definite kernels on a set X and α, β two positive scalars. Show that $\alpha K_1 + \beta K_2$ is positive definite, and describe its RKHS.

1. Because $(x, y) \mapsto xy$ is a p.d. kernel and 1 is a positive scalar, hence $(x, y) \mapsto xy + 1$ is also a p.d. kernel. Then, K_1 is a p.d. kernel, as the square of a p.d. kernel. Let's now describe its (unique) RKHS.

We already know that its RKHS contains linear combinations of the function :

$$K_x : t \mapsto K_1(x, t)$$

A function f in this RKHS can be written as :

$$f(x) = \sum_i \alpha_i K_1(x_i, x) = \sum_i \alpha_i x_i^2 x^2 + 2 \sum_i \alpha_i x_i x + \sum_i \alpha_i$$

Let's define $a = \sum_i \alpha_i x_i^2$, $b = \sum_i \alpha_i x_i$ and $c = \sum_i \alpha_i$. We have then :

$$f(x) = ax^2 + bx + c$$

Hence, our candidate RKHS H will be the set of polynomial function of order 2 :

$$f_{a,b,c}(x) = ax^2 + bx + c$$

endowed with the inner-product $\langle f_{a,b,c}, f_{a',b',c'} \rangle_H = aa' + bb' + cc' = \langle A, A' \rangle_{\mathbb{R}^3}$ where A is the vector in $\mathbb{R}^3$ composed of a, b and c and simlarly for A'.

Hence, each K_x can be seen in the RKHS as $f_{x^2,x,1}$ and therefore we have the reproducive property :

$$f_{a,b,c}(x) = ax^2 + bx + c = \langle f_{a,b,c}, K_x \rangle_H$$

Furthermore, the set of polynomial functions of order 2 is a vector space of finite dimension, hence it is complete.

Hence, the RKHS of K_1 is the set of polynomial functions of order 2, assigned to its inner product defined before.

Now, let's focus on K_2. In fact, this is not a p.d. kernel. For example, if we take $x = 1$, $y = 0$, $\alpha_1 = 1$, $\alpha_2 = -1$, we have $\alpha_1^2 K(x,x) + \alpha_2 K(y,y) + 2\alpha_1\alpha_2 K(x,y) = 0 + 1 - 2 = -1$. Hence, this is not a p.d. kernel and there is no associated RKHS.

However, if we look at the sum $K_1 + K_2$ we obtain : $K(x,y) = x^2y^2 + 1$, and here K is a p.d. kernel (with the same arguments as previously). So, let's find its (unique) RKHS.

We already know that its RKHS contains linear combinations of the function :

$$K_x : t \mapsto K(x,t)$$

A function f in this RKHS can be written as :

$$f(x) = \sum_i \alpha_i K(x_i, x) = \sum_i \alpha_i x_i^2 x^2 + \sum_i \alpha_i$$

Let's define $a = \sum_i \alpha_i x_i^2$, and $b = \sum_i \alpha_i$. We have then :

$$f(x) = ax^2 + b$$

Hence, our candidate RKHS H will be the set of polynomial function of order 2 with no first order component :

$$f_{a,b}(x) = ax^2 + b$$

endowed with the inner-product $\langle f_{a,b}, f_{a',b'} \rangle_H = aa' + bb' = \langle A, A' \rangle_{\mathbb{R}^2}$ where A is the vector in $\mathbb{R}^2$ composed of a and b, and similarly for A'.

Hence, each K_x can be seen in the RKHS as $f_{x^2,1}$ and therefore we have the reproducive property :

$$f_{a,b}(x) = ax^2 + b = \langle f_{a,b}, K_x \rangle_H$$

Furthermore, the set of polynomial functions of order 2 with no first order component is a vector space of finite dimension, hence it is complete.

Hence, the RKHS of $K_1 + K_2$ is the set of polynomial functions of order 2 with no first order component, assigned to its inner product defined before.

2. If K_1 is a p.d. kenrel and α is a positive scalar, then αK_1 is a p.d. kernel. Similarly, βK_2 with β a positive scalar, is a p.d. kernel. Hence, as the sum of 2 p.d. kernels is still a p.d. kernel, $K_1 + K_2$ is p.d. kernel. Let's find now its (unique) RKHS.

   First, the RKHS contains all the following functions :

   $$f(.) = \sum_i K_1(x_i, .) + K_2(x_i, .)$$

   Hence, the RKHS contains all the functions $f_1 + f_2$ where $f_1 \in H_1$ and $f_2 \in H_2$. The set $H_1 + H_1$ composed of such functions is a complete space, as it is the sum of 2 complete spaces (any cauchy sequence in H_1 converges in H_1, any cauchy sequence in H_2 converges in H_2, hence any cauchy sequence in $H_1 + H_2$ converges in $H_1 + H_2$). Hence, we have to find an inner product to describe the RKHS.
   Let's write the RKHS as $Span(\alpha K_1(x,.) + K_2(x,.)$. Then, any functions f and g can be written as :

   $$f(x) = \sum_i \alpha_i K_1(x_i, x) + \beta_i K_2(x_i, x)$$

   $$g(x) = \sum_i \alpha_i' K_1(x_i, x) + \beta_i' K_2(x_i, x)$$

   Let's define the inner product between f and g as :

$$\langle f, g\rangle_H = \sum_{i,j} \frac{\alpha_i}{\sqrt{\alpha}} \frac{\alpha'_i}{\sqrt{\alpha}} K_1(x_i, x_j) + \frac{\beta_i}{\sqrt{\beta}} \frac{\beta'_i}{\sqrt{\beta}} K_2(x_i, x_j)$$

We check that this is an inner product, now let us check the reproducing property. We have :

$$\langle f, \alpha K_1(x,.)+\beta K_2(x,.)\rangle_H = \langle \sum_i \alpha_i K_1(x_i,.)+\beta_i K_2(x_i,.), \alpha K_1(x,.)+\beta K_2(x,.)\rangle_H = \sum_i \alpha_i K_1(x_i,x)+\beta_i K_2(x_i,x) = f(x)$$

Hence, the reproducive property holds.

All in all, we have found the (unique) Hilbert space associated to $\alpha K_1 + \beta K_2$, wioth its associated inner product.

3 Uniqueness of the RKHS

Prove that if $K : XxX$ is a positive definite function, then it is the r.k. of a unique RKHS. To prove it, you can consider two possible RKHS H and H', and show that (i) they contain the same elements and (ii) their inner products are the same. Hint : consider the linear space spanned by the functions $K_x : t \mapsto K(x,t)$, and use the fact that a linear subspace F of a Hilbert space H is dense in H if and only if 0 is the only vector orthogonal to all vectors in F.

Suppose there are 2 RKHS for the p.d. kernel K, H_1 and H_2. Let H_0 the space composed of linear combinations of functions $K_x(x,.)$. First, we know that $H_0 \in H_1$ and $H_0 \in H_2$.

Now, let us show that H_0 is dense in H_1, by using the hint given : H_0 is dense in H_1 if and only if the vector orthogonal to all elements in H_0 is the null vector.

Let f a vector in H_0 orthogonal to all elements in H_0. Hence we have :

$$\forall g \in H_0, \langle f, g\rangle_{H_0} = 0$$

Then, we have :

$$\forall x, \langle f, K(x,.)\rangle_{H_0} = 0$$

By the reproducive property, this implies that :

$$\forall x, f(x) = 0$$

Hence, H_0 is dense in H_1. We prove also similarly that H_0 is dense in H_2.

Then, we show that both norms coincide on H_0 :

$$\forall f, g \in H_0, \langle f, g\rangle_{H_1} = \langle \sum_i \alpha_i K(x_i,.), \sum_j \beta_j K(x_j,.)\rangle_{H_1} = \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) = \langle \sum_i \alpha_i K(x_i,.), \sum_j \beta_j K(x_j,.)\rangle_{H_2}$$

Now, let f a function in H_1. Because H_0 is dense in H_1, there exists a sequence of functions in H_0 which converges to f. Because the elements of this sequence are also in H_2, there exists a sub sequence which converges in H_2. However, the limit is unique so the limit of the sub sequence in H_2 is the same as the limit of the initial sequence in H_1, which is f. Hence, $F \in H_2$. We have proven that elements from H_1 are also in H_2. We prove also the opposite side. Hence the elements in H_1 and H_2 are the same.

Now, we prove that the inner product are the same. Let $f \in H_1$ and $g \in H_2$. Let (f_n) and (g_n) the sequences of H_0 converging to f and g respectively. Then we have :

$$\langle f, g\rangle_{H_1} = \langle lim_n f_n, lim_m g_m\rangle_{H_1} = lim_n lim_m \langle f_n, g_m\rangle_{H_1} = lim_n lim_m \langle f_n, g_m\rangle_{H_0} = lim_n lim_m \langle f_n, g_m\rangle_{H_2} = \langle lim_n f_n, lim_m g_m\rangle_{H_2} =$$

Hence, the inner products are the same.

Therefore, we have $H_1 = H_2$.