Math of ML : Exercises 4 *

October 6, 2025

First some warm-up exercises on the definitions of positive definite kernel and of RKHS.

Exercise 1 (Cauchy-Schwarz inequality). *Let $k : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ be a positive definite kernel (recall Definition 3.1 in Lecture 4). Prove that for any $x, y \in \mathcal{X}$ we have $k(x, y)^2 \leq k(x, x)k(y, y)$. Deduce that $\sup_{x,y} k(x, y) = \sup_x k(x, x)$ (it may be $+\infty$).*

Solution 1. *Denote $x_1 = x$ and $x_2 = y$. By the definition of positive definite kernel, the matrix $K \in \mathbb{R}^{2\times 2}$ given by $K_{ij} = k(x_i, x_j)$ $(i, j \in \{1, 2\})$ is positive semi-definite. Hence, the determinant of K is non-negative, which proves the desired inequality.*

We have

$$\forall x,\ k(x, x) \leq \sup_y k(x, y) \leq \sqrt{k(x, x)}\sqrt{\sup_y k(y, y)}.$$

Taking the supremum over x yields $\sup_{x,y} k(x, y) = \sup_x k(x, x)$.

Exercise 2 (Bounded kernels). *Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive definite kernel. Suppose that there exists some constant $b > 0$ such that for any $x \in \mathcal{X}$ we have $k(x, x) \leq b^2$. Let $\mathcal{H}$ be the RKHS associated to the kernel k and let $f \in \{g \in \mathcal{H} : \|g\|_{\mathcal{H}} \leq 1\}$ be a function in the unit ball of $\mathcal{H}$ (recall Definition 3.3 in Lecture 4). Prove that $\sup_{x\in\mathcal{X}} |f(x)| \leq b$.*

Solution 2. *For any $x \in \mathcal{X}$, we have*

$$\begin{aligned}|f(x)| &= |\langle f, k(x, \cdot)\rangle_{\mathcal{H}}| \leq \|f\|_{\mathcal{H}}\|k(x, \cdot)\|_{\mathcal{H}} \\ &\leq \|k(x, \cdot)\|_{\mathcal{H}} = \sqrt{\langle k(x, \cdot), k(x, \cdot)\rangle_{\mathcal{H}}} \\ &= \sqrt{k(x, x)} \leq b.\end{aligned}$$

The next exercise lets you construct kernels from other kernels. It can be instructive to also ask yourself what the corresponding RKHS's are (there is not always a simple answer).

Exercise 3 (Calculus of positive definite kernels). *Let $k_1, k_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be two positive definite kernels. Show that the following functions k are also positive definite kernels.*

1. $k(x, y) = k_1(x, y) + k_2(x, y)$.
2. $k(x, y) = ak_1(x, y)$, *where* $a > 0$.
3. $k(x, y) = k_1(x, y)k_2(x, y)$.
4. $k(x, y) = p(k_1(x, y))$, *where p is a polynomial with positive coefficients.*
5. $k(x, y) = \lim_{l\to\infty} k_l(x, y)$ *(provided that the limit exists), where $k_1, k_2, k_3 \ldots$. is a sequence of positive definite kernels over $\mathcal{X}$.*

*Lénaïc Chizat EPFL `lenaic.chizat@epfl.ch`

6. $k(x, y) = \exp(k_1(x, y))$.
7. $k(x, y) = k_1(x, y)/\sqrt{k_1(x, x)k_1(y, y)}$ (you may assume $k_1(x, x) > 0$ for any $x \in \mathcal{X}$).
8. $k(x, y) = f(x) \cdot f(y)$ where $f : \mathcal{X} \to \mathbb{R}$ is any function.
9. $k(x, y) = h(\psi(x), \psi(y))$, where $h : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is a p.d. kernel and $\psi : \mathcal{X} \to \mathcal{Y}$ is arbitrary.

Which of the following functions are positive definite kernels over $\mathcal{X} = \mathbb{R}^d$?

- $k(x, y) = \|x - y\|$.
- $k(x, y) = \|x - y\|^2$.
- $k(x, y) = x^\top By$, where $B \in \mathbb{R}^{d\times d}$ is a symmetric positive-semi-definite matrix.
- $k(x, y) = \exp(-\frac{\|x-y\|^2}{2\sigma^2})$, where $\sigma > 0$.

Solution 3. *By Aronszajn's theorem (Theorem 3.2, Lecture 4), for positive definite kernels k_i, there exist Hilbert spaces $\mathcal{H}_i$ such that $k_i(x, y) = \langle \phi_i(x), \phi_i(y)\rangle$ for feature maps $\phi_i : \mathcal{X} \to \mathcal{H}_i$. Now let $\{x_1, \ldots, x_m\} \subseteq \mathcal{X}$ be an arbitrary collection of points to which we associate kernel matrices $K = (k(x_i, x_j))_{i,j=1}^n$ and $K_l = (k_l(x_i, x_j))_{i,j}^n$.*

1. Let $\alpha \in \mathbb{R}^m$ be arbitrary; then $\alpha^\top K\alpha = \alpha^\top(K_1 + K_2)\alpha = \alpha^\top(K_1)\alpha + \alpha^\top(K_2)\alpha \geq 0$.
2. $K = aK_1$ is positive semi-definite because K_1 is positive semi-definite and $a > 0$.
3. We have
$$\begin{aligned} k(x, y) &= k_1(x, y)k_2(x, y) \\ &= \left(\sum_i (\phi_1(x))_i(\phi_1(y))_i\right)\left(\sum_j (\phi_2(x))_j(\phi_2(y))_j\right) \\ &= \sum_{i,j} (\phi_1(x))_i(\phi_2(x))_j \cdot (\phi_1(y))_i(\phi_2(y))_j \\ &= \langle \phi(x), \phi(y)\rangle, \end{aligned}$$
where $\phi(x) = ((\phi_1(x)_i(\phi_2(x))_j)_{ij}$. Hence, by Aronszajn's theorem, k is a positive definite kernel.
4. The case $k(x, y) = p(k_1(x, y))$, where p is a polynomial with positive coefficients and k_1 is a p.d. kernel, follows from the combination of the three previous items.
5. Let $\alpha \in \mathbb{R}^m$ be arbitrary. Since k_l (for $l = 1, 2, \ldots$) is a positive definite kernel, we have $\alpha^\top K_l\alpha \geq 0$. It follows that $\alpha^\top K\alpha = \lim_{l\to\infty} \alpha^\top K_l\alpha \geq 0$, which shows that k is a positive definite kernel.
6. The case $k(x, y) = \exp(k_1(x, y))$ follows by items 4. and 5. by considering $h_l(x, y) = p_l(k_1(x, y))$ where $p_l(X) = \sum_{s=1}^l \frac{1}{s!}X^s$.
7. We have
$$k(x, y) = \frac{\langle \phi'(x), \phi'(y)\rangle}{\|\phi'(x)\| \cdot \|\phi'(y)\|} = \langle \phi(x), \phi(y)\rangle$$
for the feature map $\phi : \mathcal{X} \to \mathcal{H}'$ defined by $\phi(x) = \phi'(x)/\|\phi'(x)\|$. Hence, k is a positive definite kernel.

   Remark: the result of this question is also valid without the positivity assumption $k(x, x) > 0$.

8. *The case of* $k(x,y) = f(x)\cdot f(y)$ *follows directly from Aroszajn theorem.*

9. *Let* ϕ *be a feature map for* h *as in Aronszajn's theorem. Then* $k(x,y) = h(\psi(x),\psi(y)) = \langle(\phi\circ\psi)(x),(\phi\circ\psi)(y)\rangle$*, so* k *is a p.d. kernel.*

For the specific examples (or counter-examples):

- *Let* $x_1 = 0$ *and let* x_2 *be any unit vector. Then the* 2×2 *matrix* $K_{i,j} = k(x_i,x_j)$ *is not positive semi-definite. Hence,* k *is not a p.d. kernel.*

- *The same counter-example as just above shows that this* k *is not a p.d. kernel either.*

- *Consider* $B = \sum_{i=1}^d \lambda_i u_i u_i^\top$ *the eigenvalue decomposition of* B*. Then we have* $k(x,y) = \sum_{i=1}^d \lambda_i \varphi_i(x)\varphi_i(y)$ *where* $\varphi_i(z) = z^\top u_i$*, so by items 8., 2., and 1. above,* k *is a p.d. kernel.*

- *By item 6., we have that* $h:(x,y)\mapsto \exp(x^\top y/\sigma^2)$ *is a p.d. kernel. So by applying item 7., we get that*

$$k(x,y) = \frac{k(x,y)}{\sqrt{k(x,x)k(y,y)}} = \frac{e^{x^\top y/\sigma^2}}{e^{\|x\|^2/(2\sigma)}\cdot e^{\|y\|^2(/2\sigma)}} = e^{\|x-y\|^2/(2\sigma)}$$

  is indeed a p.d. kernel. It is called the (Gaussian) radial basis function kernel, or RBF kernel.

In this next exercise you characterize all translation-invariant kernels on the 1-D torus. (Translation-invariant kernels on $\mathbb{R}^d$ will be discussed in the next lecture.)

Exercise 4 (Translation-invariant kernels on $[0,1]$ [Bach, 2024, Sec. 7.3.2])**.**

1. *Let* q *be a continuous* 1*-periodic function and* $k(x,y) = q(x-y)$*. Show that* k *is a positive definite kernel on* $[0,1]$ *if and only if all of the Fourier coefficients of* q *are non-negative. Express the associated RKHS in terms of the Fourier coefficients of* q*.*

   Hint/Reminder: The Fourier series decomposition of a function $f\in L_2([0,1])$ *is* $f(x) = \sum_{m\in\mathbb{Z}} e^{2im\pi x}\hat{f}_m$*, and the Fourier coefficients are given by* $\hat{f}_m = \int_0^1 f(x)e^{-2im\pi x}\mathrm{d}x \in \mathbb{C}$*.*

2. *Using question 1., show that the set* $\left\{ f:[0,1]\to\mathbb{R} \text{ s.t. } \left(\int_0^1 f\right)^2 + \int_0^1 |f'(x)|^2\mathrm{d}x < \infty \right\}$ *equipped with the inner product*

$$\langle f,g\rangle = \left(\int_0^1 f\right)\left(\int_0^1 g\right) + \frac{1}{(2\pi)^2}\int_0^1 f'(x)g'(x)\mathrm{d}x$$

   is a RKHS and identify the corresponding kernel function (you may leave it in the form of a trigonometric series, $\sum_n A_n\cos(nX) + B_n\sin(nX)$*).*

Solution 4. *See [Bach, 2024, Sec. 7.3.2].*

References

Francis Bach. *Learning theory from first principles.* MIT press, 2024. URL https://www.di.ens.fr/~fbach/ltfp_book.pdf.