Homework 2
Due November 24th

1 Combination rules for kernels

Consider a set $\mathcal{X}$ and two positive definite (p.d.) kernels $K_1, K_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

1. For all scalars $\alpha, \beta \geq 0$, show that the sum kernel $\alpha K_1 + \beta K_2$ is p.d.
2. Show that the product kernel $(x, y) \mapsto K_1(x, y)K_2(x, y)$ is p.d. (Be careful, this is a pointwise multiplication, not a matrix multiplication)
   - Tip1: It is sufficient to show that for any finite set $x_1, \ldots, x_n$ in $\mathcal{X}$, the corresponding $n \times n$ kernel matrix is positive semi-definite. Remember that a positive semi-definite matrix A in $\mathbb{R}^{n\times n}$ can be factorized into a product $Z^\top Z$ with Z in $\mathbb{R}^{n\times n}$.
   - Tip2: remember the proof for showing that the polynomial kernel $(x, y) \mapsto (x^\top y)^2$ is p.d. with the "trace" trick. This was indeed a product kernel with $K_1(x, y) = K_2(x, y) = (x^\top y)$ and $\mathcal{X} = \mathbb{R}^p$.
3. Given a sequence $(K_n)_{n\geq 0}$ of p.d. kernels such that for all x, y in $\mathcal{X}$, $K_n(x, y)$ converges to a value $K(x, y)$ in $\mathbb{R}$ (pointwise convergence). Show that K is a p.d. kernel.
4. Show that the kernel $K(x, y) = 1/(1 - x^\top y)$ with $\mathcal{X} = \{x \in \mathbb{R}^p : \|x\|_2 < 1\}$ is p.d.

2 The kernel $K(x, y) = \min(x, y)$ with $\mathcal{X} = [0, 1]$

A function $f : \mathcal{X} \to \mathbb{R}$ is said to be "absolutely continuous" if the function is differentiable almost everywhere with f' integrable and $f(x) = f(0) + \int_{t=0}^{x} f'(t)dt$ for all x in $\mathcal{X}$. Surprisingly, we will see that this concept is closely related to the min kernel.

1. Show that the kernel K is p.d.
   Tip: try to write $\min(x, y)$ as an integral.
2. Show that the functional space $\mathcal{H}$ below is Hilbertian (you need to define an appropriate inner-product):

$$\mathcal{H} = \left\{ f : \mathcal{X} \to \mathbb{R} \text{ such that } f \text{ is absolutely continuous, } f' \in L^2(\mathcal{X}) \text{ and } f(0) = 0 \right\},$$

where

$$L^2(\mathcal{X}) = \left\{ f : \mathcal{X} \to \mathbb{R} \text{ such that } \int_{t \in \mathcal{X}} f(t)^2 dt < +\infty \right\}$$

is a Hilbert space with inner product $\langle f, g \rangle_{L^2(\mathcal{X})} = \int_{t\in\mathcal{X}} f(t)g(t)dt$. [1]

3. Show that K is the reproducing kernel associated to $\mathcal{H}$ for an appropriate inner-product.

Remark: even though, we do not precise it in the text for simplicity, we always consider Lebesgue measurable functions.

3 Shift-invariant kernels

A large class of kernels used in machine learning have a property of shift-invariance. We consider in this exercise kernels defined on the set $\mathcal{X} = \mathbb{R}$, but note that extensions to other sets, including multi-dimensional spaces is possible. A kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called shift-invariant when there exists a symmetric function $\kappa : \mathcal{X} \to \mathbb{R}$ such that for all x, y in $\mathcal{X}$,

$$K(x, y) = \kappa(x - y),$$

where symmetric means that $\kappa(-t) = \kappa(t)$ for all t in $\mathbb{R}$. We consider from now on such a function κ and also assume that κ is continuous and in $L^1(\mathbb{R})$ where

$$L^1(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} \text{ such that } \int_{t \in \mathbb{R}} |f(t)| dt < +\infty \right\}.$$

We will show in this exercise that describing the function space associated to such kernels involves characterizing the Fourier transform of κ, denoted $\hat{\kappa} : \mathbb{R} \to \mathbb{C}$, and defined as

$$\hat{\kappa} : \omega \mapsto \int_{t \in \mathbb{R}} e^{-it\omega} \kappa(t) dt,$$

where i is the imaginary unit. We assume a few preliminary results to be true without proving them.

- if $\hat{\kappa}$ is in $L^1(\mathbb{R})$, then for all x in $\mathbb{R}$, we have the inverse Fourier transform formula

$$\kappa(x) = \frac{1}{2\pi} \int_{\omega \in \mathbb{R}} e^{ix\omega} \hat{\kappa}(\omega) d\omega.$$

- the translation invariant kernel K above is positive definite if and only if $\hat{\kappa}$ is real-valued, symmetric, and non-negative. (This is a consequence of Bochner theorem[2]).

[1] Note that to be rigorous, we should remark that $L^2(\mathcal{X})$ is a Hilbert space only when two functions that are equal almost everywhere are considered to be the same element of $L^2(\mathcal{X})$. This can be formalized by defining a quotient space with an equivalence class, or by using the concept of distribution. To simplify, we will omit this subtlety, since it will not affect any result of the homework.

[2] Note the here as well, we should in fact use the concept of distribution to be perfectly rigorous.

- the subset $\mathcal{H}$ of $L^2(\mathbb{R})$ consisting of integrable and continuous functions f such that

  $$\int_{\omega \in \mathbb{R}} \frac{|\hat{f}(\omega)|^2}{\hat{\kappa}(\omega)} d\omega < +\infty,$$

  endowed with the inner-product

  $$\langle f, g \rangle_{\mathcal{H}} = \frac{1}{2\pi} \int_{\omega \in \mathbb{R}} \frac{\hat{f}(\omega)\hat{g}(\omega)^*}{\hat{\kappa}(\omega)} d\omega,$$

  is a Hilbert space, where * denotes the complex conjugate.

The questions of this exercise are now the following:

1. Show that $\mathcal{H}$ is the RKHS associated to K.

2. One of the most widely used kernel in machine learning is the Gaussian kernel:

   $$K(x, y) = e^{-\frac{1}{2\sigma^2}(x-y)^2}.$$

   Show that it is p.d. and describe its feature space $\mathcal{H}$.

   Remark: computing the Fourier transform of a Gaussian is hard to do if you have not seen it before. You may look into classical maths textbooks to find the trick explaining how to do it.

3. Do the same for the Laplace kernel:

   $$K(x, y) = \frac{1}{2} e^{-\gamma |x-y|}.$$

   Tip: integration by part will be helpful.