Homework
Due February 7th

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)
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 e^{K_1} is p.d.

2 Positive definite kernels

Which of these kernels are positive definite? You need to provide proofs for all cases.

- $K(x, y) = 1/(1 - xy)$ with $\mathcal{X} = (-1, 1)$.
- $K(x, y) = 2^{xy}$ with $\mathcal{X} = \mathbb{N}$.
- $K(x, y) = \log(1 + xy)$ with $\mathcal{X} = \mathbb{R}_+$.
- $K(x, y) = e^{-(x-y)^2}$ with $\mathcal{X} = \mathbb{R}$.
- $K(x, y) = \cos(x + y)$ with $\mathcal{X} = \mathbb{R}$.
- $K(x, y) = \cos(x - y)$ with $\mathcal{X} = \mathbb{R}$.
- $K(x, y) = \min(x, y)$ with $\mathcal{X} = \mathbb{R}_+$.
- $K(x, y) = \max(x, y)$ with $\mathcal{X} = \mathbb{R}_+$.
- $K(x, y) = \min(x, y)/\max(x, y)$ with $\mathcal{X} = \mathbb{R}_+$.
- $K(x, y) = GCD(x, y)$ (greatest common divisor) with $\mathcal{X} = \mathbb{N}$.
- $K(x, y) = LCM(x, y)$ (least common multiple) with $\mathcal{X} = \mathbb{N}$.
- $K(x, y) = GCD(x, y)/LCM(x, y)$ (least common multiple) with $\mathcal{X} = \mathbb{N}$.

3 Covariance Operators in RKHS

Given two sets of real numbers $X = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $Y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, the covariance between X and Y is defined as

$$cov_n(X, Y) = \mathbb{E}_n(XY) - \mathbb{E}_n(X)\mathbb{E}_n(Y)\,,$$

where $\mathbb{E}_n(U) = (\sum_{i=1}^n u_i)/n$. The covariance is useful to detect linear relationships between X and Y. In order to extend this measure to potential nonlinear relationships between X and Y, we consider the following criterion:

$$C_n^K(X, Y) = \max_{f,g \in \mathcal{B}_K} cov_n(f(X), g(Y))\,,$$

where K is a positive definite kernel on $\mathbb{R}$, $\mathcal{B}_K$ is the unit ball of the RKHS of K, and $f(U) = (f(u_1), \ldots, f(u_n))$ for a vector $U = (u_1, \ldots, u_n)$.
1. Express simply $C_n^K(X, Y)$ for the linear kernel $K(a, b) = ab$.
2. For a general kernel K, express $C_n^K(X, Y)$ in terms of the Gram matrices of X and Y.

4 Some upper bounds for learning theory

Let K be a positive definite kernel on a measurable set $\mathcal{X}$, $(\mathcal{H}_K, \|\,.\,\|_{\mathcal{H}_K})$ denote the corresponding reproducing kernel Hilbert space, $\lambda > 0$, and $\phi : \mathbb{R} \to \mathbb{R}$ a function. We assume that:

$$\kappa = \sup_{\mathbf{x} \in \mathcal{X}} K(\mathbf{x}, \mathbf{x}) < +\infty\,,$$

and we note $B_R = \{f \in \mathcal{H}_K, \| f \|_{\mathcal{H}_K} \leq R\}$. Let us define, for all $f \in \mathcal{H}$ and $\mathbf{x} \in \mathcal{X}$,

$$R_\phi(f, \mathbf{x}) = \phi(f(\mathbf{x})) + \lambda \| f \|^2_{\mathcal{H}_K}\,.$$

1. ϕ is said to be Lipschitz if there exists a constant $L > 0$ such that, for all $u, v \in \mathbb{R}$, $|\,\phi(u) - \phi(v)\,| \leq L\,|\,u - v\,|$. Show that, in that case, there exists a constant C_1 to be determined such that, for all $\mathbf{x} \in \mathcal{X}$ and $f, g \in B_R$:

$$|\,R_\phi(f, \mathbf{x}) - R_\phi(g, \mathbf{x})\,| \leq C_1 \| f - g \|_{\mathcal{H}_K}\,.$$

2. ϕ is said to be convex if for all $u, v \in \mathbb{R}$ and $t \in [0, 1]$, $\phi(tu + (1-t)v) \leq t\phi(u) + (1 - t)\phi(v)$. We assume that ϕ is convex, and that for all $\mathbf{x} \in \mathcal{X}$, there exists $f_\mathbf{x} \in \mathcal{H}$ which minimizes $f \mapsto R_\phi(f, \mathbf{x})$. Show that there exists a constant $C_2 > 0$ to be determined, such that:

$$\psi(f, \mathbf{x}) \triangleq R_\phi(f, \mathbf{x}) - R_\phi(f_\mathbf{x}, \mathbf{x}) \geq C_2 \| f - f_\mathbf{x} \|^2_{\mathcal{H}_K}\,.$$

3. Under the hypothesis of questions **2.1** and **2.2**, show that there exists a constant C, to be determined, such that if X is a random variable with values in $\mathcal{X}$, then:

$$\forall f \in B_R, \quad \mathbb{E}\psi(f, X)^2 \leq C\mathbb{E}\psi(f, X)\,.$$