Advanced learning models - 2nd homework

Due February 16th 2020

Exercice 1. Some kernels...
Show that the following kernels are positive definite:

1. On $\mathcal{X} = \mathbb{R}$:
$$\forall x, y \in \mathbb{R}\,, \quad K(x,y) = \cos(x-y)\,.$$
2. On $\mathcal{X} = \{x \in \mathbb{R}^p : \|x\|_2 < 1\}$:
$$\forall x, y \in \mathcal{X}\,, \quad K(x,y) = 1/(1 - x^\top y)\,.$$
3. On $\mathcal{X} = \mathbb{N}$:
$$\forall x, y \in \mathbb{N}\,, \quad K(x,y) = (-1)^{x+y}\,.$$
4. On $\mathcal{X} = \mathbb{R}^n$:
$$\forall x, y \in \mathcal{X}\,, \quad K(x,y) = \pi - \arccos\left(\frac{x^\top y}{\|\, x \,\|\|\, y \,\|}\right)\,.$$

Exercice 2. Dual of the SVM with intercept.
We recall the primal formulation of SVMs seen in the class.

$$\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i f(x_i)) + \lambda \|f\|_{\mathcal{H}}^2,$$

and its dual formulation.

$$\max_{\alpha \in \mathbb{R}^n} 2\alpha^\top \mathbf{y} - \alpha^\top \mathbf{K} \alpha \quad \text{such that} \quad 0 \leq y_i \alpha_i \leq \frac{1}{2\lambda n}, \quad \text{for all } i. \tag{1}$$

Consider the primal formulation of SVMs with intercept

$$\min_{f \in \mathcal{H}, b \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i (f(x_i) + b)) + \lambda \|f\|_{\mathcal{H}}^2,$$

Can we still apply the representer theorem? Why? Derive the corresponding dual formulation by using Lagrangian duality. Provide a formulation in terms of α in $\mathbb{R}^n$ as in (1).

Exercice 3. Kernels encoding equivalence classes.
Consider a similarity measure $K : \mathcal{X} \times \mathcal{X} \to \{0,1\}$ with $K(x,x) = 1$ for all x in $\mathcal{X}$. Prove that K is p.d. if and only if, for all x, x', x'' in $\mathcal{X}$,

- $K(x,x') = 1 \Leftrightarrow K(x',x) = 1$, and
- $K(x,x') = K(x',x'') = 1 \Rightarrow K(x,x'') = 1$.

Exercice 4. COCO
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

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

where $\mathbf{E}_n(U) = (\sum_{i=1}^n u_i)/n$. 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.

Exercice 5. RKHS

1. Let K_1 and K_2 be two positive definite kernels on a set $\mathcal{X}$, and α, β two positive scalars. Show that $\alpha K_1 + \beta K_2$ is positive definite, and describe its RKHS.
2. Let $\mathcal{X}$ be a set and $\mathcal{F}$ be a Hilbert space. Let $\Psi : \mathcal{X} \to \mathcal{F}$, and $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be:

$$\forall x, x' \in \mathcal{X}, \quad K(x,x') = \langle \Psi(x), \Psi(x') \rangle_{\mathcal{H}}\ .$$

   Show that K is a positive definite kernel on $\mathcal{X}$, and describe its RKHS.
3. Prove that for any p.d. kernel K on a space $\mathcal{X}$, a function $f : \mathcal{X} \to \mathbb{R}$ belongs to the RKHS $\mathcal{H}$ with kernel K if and only if there exists $\lambda > 0$ such that $K(\mathbf{x}, \mathbf{x}') - \lambda f(\mathbf{x}) f(\mathbf{x}')$ is p.d.