Kernel methods for machine learning
Homework 1

Med Chiheb Yaakoubi

Exercice 1: Kernels

- **K(x,x')=min(x,x')**
  Let $n \in \mathbb{N}$, $x = (x_i)_i \in \mathbb{R}_+^n$, $a = (a_i)_i \in \mathbb{R}^n$ and finally $A_k = (\mathbf{1}_{i\geq k}\mathbf{1}_{j\geq k})_{i,j\in n}$
  we can see that:
  $$a^T A_k a = \sum_{i=k}^{n}\sum_{j=k}^{n} a_i a_j = (\sum_{i=k}^{n} a_i)^2 \geq 0$$
  But if we take $0 \leq x_1 \leq ... \leq x_n$ without loss of generality we get that the Kernel matrix K_x verifies:
  $$K_x = \sum_{i=0}^{n}(x_{i+1} - x_i)A_i$$
  such that $x_0 = 0$ and since this is a linear combination of positive semidefinite matrices with positive coefficients then this gives us that K_x is also positive semidefinite. Given the symmetry of K we get that **K:(x,x')$\in \mathbb{R}_+^2 \mapsto$ min(x,x') is p.d**

- **K(x,x')=max(x,x')**
  If we suppose that K is a p.d then we get that :
  $$K' : (x, x') \in {\mathbb{R}_+}^2 \mapsto \max(x, x') + \min(x, x') = x' + x$$
  is also a p.d but if we take $x = (0, 1)$, $a = (2, -1)$ we get:
  $$a^T K'_x a = \begin{pmatrix} 2 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = -3$$
  This gives us an absurd result hence, **max(x,x') is not a pd**

- **K(x,y)=min(f(x)g(y),f(y)g(x))**

Exercice 2: Non-expansiveness of the Guassian kernel

We know that the RKHS mapping verifies $\varphi(x) = K_x$, hence we get:

$$\begin{aligned}
||\varphi(x) - \varphi(x')||^2_{\mathcal{H}} &=< \varphi(x) - \varphi(x'), \varphi(x) - \varphi(x') >_{\mathcal{H}} \\
&=< K_x - K'_x, K_x - K'_x >_{\mathcal{H}} \\
&= K(x,x) + K(x',x') - 2K(x,x') \\
&= 2(1 - \exp(-\frac{\alpha}{2}||x - x'||^2)) \\
&\leq \alpha||x - x'||^2
\end{aligned}$$

Exercice 3: RKHS

- If we take $x \in \mathcal{X}^n$ for $n \in \mathbb{N}$ then we have that the kernel matrices $K_1(x)$ and $K_2(x)$ are positive semidefinite which gives us that $\alpha K_1(x) + \beta K_2(x)$ is also positive semidefinite hence $\alpha K_1 + \beta K_2$ is p.d. with it's RKHS being the closur of the span of the family

  $$(\alpha K_1(x,.) + \beta K_2(x,.))_x = (t \in \mathcal{X} \mapsto \alpha K_1(x,t) + \beta K_2(x,t))_x$$

  and it's dot product being the following:

  $$< \sum_i a_i(\alpha K_1(x_i,.) + \beta K_2(x_i,.)), \sum_i b_i(\alpha K_1(y_i,.) + \beta K_2(y_i,.)) >_{\mathcal{H}} = \sum_{i,j} a_i b_i(\alpha K_1(x_i,y_j) + \beta K_2(x_i,y_i))$$

- ($\implies$) for $K(x,x') - \lambda f(x)f(x')$ to be p.d we have to get that :

  $$\sum_{i,j} a_i a_j K(x_i,x_j) - \lambda f(x_i)f(x_j) \geq 0$$

  $$\sum_{i,j} a_i a_j K(x_i,x_j) \geq \lambda \sum_i a_i a_j f(x_i)f(x_j)$$

  $$||\sum_i a_i K_{x_i}|| \geq \lambda < f, \sum_i a_i K_{x_i} >^2$$

  But with cauchy shwartz we get :

  $$\frac{||\sum_i a_i K_{x_i}||}{< f, \sum_i a_i K_{x_i} >^2} \geq \frac{1}{||f||^2_{\mathcal{H}}}$$

  Hence for $\lambda \in]0, \frac{1}{||f||^2_{\mathcal{H}}}[$ we get the needed result

  ($\impliedby$) Let's define $\varphi : \sum_i a_i K_{x_i} \mapsto \sum_i a_i f(x_i)$ we can see that thanks to the fact that $K(x,x') - \lambda f(x)f(x')$ is pd that φ is $\frac{1}{\sqrt{\lambda}}$-Lipschitz meaning it's continuous. This also gives us that we can define it on the whole of $\mathcal{H}$ (given also the density of the old domaian). By riez theorem we get that:

  $$\exists \tilde{f} \in \mathcal{H} \setminus \varphi(x) =< \tilde{f}, x >_{\mathcal{H}} \ \forall x \in \mathcal{H}$$

  but since

  $$f(x) = \varphi(x) =< \tilde{f}, x >_{\mathcal{H}} = \tilde{f}(x) \ \forall x \in \mathcal{X}$$

  this gives us that $f \in \mathcal{H}$