Kernel methods in machine learning — Homework 1

Hugo Cisneros

March 13th, 2019

Exercise 1. Kernels

1.1 K is symmetric since

$$\forall x, y \in \mathbb{R}^2 K(y,x) = \cos(-(x-y)) = \cos(x-y) = K(x,y)$$

and because for all sequences of (x_i) and (a_i) of $\mathbb{R}$

$$\begin{aligned}\sum_i \sum_j a_i a_j \cos(x_i - x_j) &= \sum_i \sum_j a_i a_j (\cos x_i \cos x_j + \sin x_i \sin x_j) \\ &= \left(\sum_i a_i \cos x_i\right)^2 + \left(\sum_i a_i \sin x_i\right)^2 \geq 0\end{aligned}$$

K is p.d.

1.2 K is clearly symmetric and $\forall x, y \in \mathcal{X}$,

$$\begin{aligned}K(x,y) &= \frac{1}{1 - x^T y} \\ &= \lim_{n\to\infty} \sum_{k=0}^{n} (x^T y)^k && \text{(converges because by C-S } |x^T y| \leq \|x\|_2 \|y\|_2 < 1)\end{aligned}$$

Since all powers of the linear kernel are p.d. kernels, and a sum of p.d. kernels is a p.d. kernel, the above sum is a p.d. kernel for all n. Therefore, K is p.d. as a point-wise limit of a sequence of p.d. kernels.

1.3 The kernel K is symmetric because the intersection is a commutative operation.

$$\begin{aligned}\sum_{i=1}^{n} a_i a_j (P(A_i \cap A_j) - P(A_i)P(A_j)) &= \sum_{i=1}^{n} a_i a_j (\mathbb{E}[\mathbb{1}_{A_i \cap A_j}] - \mathbb{E}[\mathbb{1}_{A_i}]\mathbb{E}[\mathbb{1}_{A_j}]) \\ &= \mathbb{E}\left[\sum_{i=1}^{n} a_i a_j \mathbb{1}_{A_i} \mathbb{1}_{A_j}\right] - \sum_{i=1}^{n} \mathbb{E}[a_i \mathbb{1}_{A_i}]\mathbb{E}[a_j \mathbb{1}_{A_j}] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}\right)^2\right] - \left(\sum_{i=1}^{n} a_i \mathbb{E}[\mathbb{1}_{A_i}]\right)^2\end{aligned}$$

By Jensen's inequality, the above quantity is positive because the function $\phi : (X_1, ..., X_n) \mapsto (\sum_{i=1}^{n} a_i X_i)^2$ is convex. Therefore, K **is p.d.**.

1.4

1.5

Exercise 2. RKHS

2.1 $\alpha K_1 + \beta K_2$ **is p.d.** as sum of p.d. kernels and because α and β are positive scalars. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be their respective RKHS.

Since, K_1 and K_2 are p.d., by Aronszajn's theorem there exist Φ_1, Φ_2 mappings from $\mathcal{X}$ to $\mathcal{H}_1$ and $\mathcal{H}_2$ such that $\forall x, y \in \mathbb{R}^2$

$$\begin{aligned}
K(x,y) &\triangleq \alpha K_1(x,y) + \beta K_2(x,y) \\
&= \langle \sqrt{\alpha}\Phi_1(x), \sqrt{\alpha}\Phi_1(y)\rangle_{\mathcal{H}_1} + \langle \sqrt{\beta}\Phi_2(x), \sqrt{\beta}\Phi_2(y)\rangle_{\mathcal{H}_2} \\
&= \langle \Phi(x), \Phi(y)\rangle_{\mathcal{H}}
\end{aligned}$$

Where $\langle ., .\rangle_{\mathcal{H}}$ is defined above and

$$\begin{aligned}
\Phi : \mathcal{X} &\to \mathcal{H} \triangleq \sqrt{\alpha}\mathcal{H}_1 + \sqrt{\beta}\mathcal{H}_2 \\
x &\mapsto \sqrt{\alpha}\Phi_1(x) + \sqrt{\beta}\Phi_2(x)
\end{aligned}$$

$\langle ., .\rangle_{\mathcal{H}}$ is clearly an inner product. And for any $(x_n)_{n\in\mathbb{N}}$ sequence of $\mathcal{H}$, we have that $\forall n \in \mathbb{N}, x_n = \sqrt{\alpha}x_{1,n} + \sqrt{\beta}x_{2,n}$ and $\forall n, m \in \mathbb{N}$

$$\begin{aligned}
\|x_n - x_m\|^2_{\mathcal{H}} &= \|\sqrt{\alpha}(x_{1,n} - x_{1,m}) + \sqrt{\beta}(x_{2,n} - x_{2,m})\|^2_{\mathcal{H}} \\
&= \langle \sqrt{\alpha}(x_{1,n} - x_{1,m}) + \sqrt{\beta}(x_{2,n} - x_{2,m}), \sqrt{\alpha}(x_{1,n} - x_{1,m}) + \sqrt{\beta}(x_{2,n} - x_{2,m})\rangle_{\mathcal{H}} \\
&= \alpha\langle x_{1,n} - x_{1,m}, x_{1,n} - x_{1,m}\rangle_{\mathcal{H}_1} + \beta\langle x_{2,n} - x_{2,m}, x_{2,n} - x_{2,m}\rangle_{\mathcal{H}_2} \qquad \text{(by definition of } \langle ., .\rangle_{\mathcal{H}}) \\
&= \alpha\|x_{1,n} - x_{1,m}\|^2_{\mathcal{H}_1} + \beta\|x_{2,n} - x_{2,m}\|^2_{\mathcal{H}_2}
\end{aligned}$$

Therefore, if (x_n) is a Cauchy sequence, $(x_{1,n})$ and $(x_{2,n})$ are also Cauchy sequences. Because $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces, the two sequences converge and (x_n) converges. $\mathcal{H}$ **is a Hilbert space**.

Let's consider the functions $K_x : t \mapsto K(x,t) = \alpha K_1(x,t) + \beta K_2(x,t)$ for $x \in \mathcal{X}$

Let, $f \in \mathcal{H}$, $x \in \mathcal{X}$. f can be written $f = \sqrt{\alpha}f_1 + \sqrt{\beta}f_2$ with $f_1 \in \mathcal{H}_1$ and $f_2 \in \mathcal{H}_2$.

$$\begin{aligned}
f(x) &= \sqrt{\alpha}f_1(x) + \sqrt{\beta}f_2(x) \\
&= \sqrt{\alpha}\langle f_1, K_{1,x}\rangle_{\mathcal{H}_1} + \sqrt{\beta}\langle f_2, K_{2,x}\rangle_{\mathcal{H}_2} \\
&= \langle f, K_x\rangle_{\mathcal{H}}
\end{aligned}$$

2.2

Exercise 3. RKHS

3.1 $\mathcal{H}$ is Hilbert:

$\mathcal{H}$ is a vector space of functions. $\langle ., .\rangle_{\mathcal{H}}$ is a symmetric bilinear form verifying $\forall f, \langle f, f\rangle_{\mathcal{H}} \geq 0$.

Since f is absolutely continuous, it has a derivative almost everywhere and the following equality holds $\forall x \in [0,1]$

$$\begin{aligned}
|f(x)|^2 &= \left| f(0) + \int_0^x f'(x)\mathrm{d}x \right|^2 \\
&= \left| \int_0^x f'(x)\mathrm{d}x \right|^2 && (f(0)=0) \\
&\leq x \cdot \int_0^x f'(u)^2 \mathrm{d}u = x \cdot \langle f, f \rangle_{\mathcal{H}}
\end{aligned}$$

$\langle f, f \rangle_{\mathcal{H}} \implies f = 0$ and $\langle ., . \rangle_{\mathcal{H}}$ is therefore and inner product. **$\mathcal{H}$ is a pre-Hilbert space with $\langle ., . \rangle_{\mathcal{H}}$ as inner product**.

Let (f_n) a Cauchy sequence of $\mathcal{H}$. (f_n') is a Cauchy sequence of $L^2([0,1])$ which is complete. Therefore it converges to a function $g \in L^2([0,1])$.

Since for all (n,m), $x \in [0,1]$, $|f_n(x) - f_m(x)|^2 \leq x \cdot \|f_n - f_m\|_{\mathcal{H}}^2$, the sequence $f_n(x)$ is Cauchy for any x and converges to a real number $f(x)$. And since $f(x) = \lim_\infty f_n(x) = \lim_\infty \int_0^x f_n'(u)\mathrm{d}u = \int_0^x g(u)\mathrm{d}u$, f is absolutely continuous and $f' = g$ almost everywhere. Moreover, $f' \in L^2([0,1])$ and $f(0) = \lim_\infty f_n(0) = 0$.

Finally, $\|f_n - f\|_{\mathcal{H}} = \|f_n' - g\|_{L^2([0,1])} \to 0$ and $f \in \mathcal{H}$. **$\mathcal{H}$ is complete, therefore $\mathcal{H}$ is a Hilbert space**.

Reproducing property:

We will now show that $\mathcal{H}$ is the RKHS with corresponding kernel $K : (x,y) \to \min(x,y)$ on $[0,1]$.

For $x \in [0,1]$, the function $K_x = \min(x, \cdot)$ is differentiable except on the singleton x which has a null measure, it is absolutely continuous. Its derivative is square integrable and $\min(x,0) = 0$, **therefore K_x is in $\mathcal{H}$ for all x**.