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 $\forall x, y \in \mathcal{X}$,

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

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.

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 $\forall x \in [0, 1]$

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

$\langle f, f \rangle_{\mathcal{H}} \implies f = 0$ 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), $x \in [0, 1]$, $|f_n(x) - f_m(x)|^2 \leq x \cdot \|f_n - f_m\|^2_{\mathcal{H}}$, 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, $\|f_n - f\|_{\mathcal{H}} = \|f'_n - g\|_{L^2([0,1])} \xrightarrow[n \to +\infty]{} 0$ 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 $[0, 1]^2$.

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.

For any $x \in [0, 1]$ and $f \in \mathcal{H}$

$$\langle f, K_x \rangle_{\mathcal{H}} = \int_0^1 f'(u) K'_x(u)\mathrm{d}u = \int_0^x f'(u)\mathrm{d}u = f(x)$$

Therefore, K **is the r.k. of the RKHS** $\mathcal{H}$.

3.2 The above demonstration also shows that $\mathcal{H}$ **is a Hilbert space** by remarking that

$$f(1) = \lim_{\infty} f_n(1) = 1$$

for (f_n) Cauchy sequence of elements of $\mathcal{H}$.

We will now show that the r.k. corresponding to this Hilbert space is $K : (x, y) \mapsto \min(x, y) - xy$.

For $x \in [0, 1]$, the function K_x has a derivative everywhere except on $\{x\}$ and its derivative is square integrable. Moreover, $K_x(0) = 0$ and $K_x(1) = 0$, therefore K_x **is in** $\mathcal{H}$.

Let $x \in [0, 1]$ and $f \in \mathcal{H}$

$$\langle f, K_x \rangle_{\mathcal{H}} = \int_0^1 f'(u) K_x'(u) \mathrm{d}u = (1 - x) \int_0^x f'(u) \mathrm{d}u - x \int_x^1 f'(u) \mathrm{d}u = f(x)$$

Therefore K **is the r.k. of the RKHS** $\mathcal{H}$.

3.3

Exercise 4. Duality

4.a The problem is a convex problem for which strong duality holds because it is strictly feasible (e.g the point $f = 0$ satisfies the inequality constraint). Therefore, there exist a dual parameter such that the problem is equivalent to

$$\min_{f \in \mathcal{H}_\mathcal{K}} \quad \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}(f(x_i)) + \lambda \|f\|_{\mathcal{H}_\mathcal{K}}$$

According to the representer theorem, the above problem has a solution of the form $\forall x \in \mathcal{X}$

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x)$$

The optimization problem can be re-written as

$$\min_{\alpha \in \mathbb{R}^n} \quad R(K\alpha) + \lambda \alpha^\top K \alpha$$

where

$$\begin{aligned} R : \mathbb{R}^n &\to \mathbb{R} \\ ([K\alpha]_1, ..., [K\alpha]_n) &\mapsto \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}([K\alpha]_i) \end{aligned}$$

4.b From the definition

$$\begin{aligned} R^*(u) &= \sup_{x \in \mathbb{R}^n} \left(x^\top u - R(x) \right) \\ &= \sup_{x \in \mathbb{R}^n} \frac{1}{n} \sum_{i=1}^{n} (n x_i u_i - \ell_{y_i}(x_i)) \\ &= \frac{1}{n} \sum_{i=1}^{n} \ell_{y_i}^*(n u_i) \end{aligned}$$

4.c The Laplacian of the problem is

$$L(u, \alpha, \mu) = R(u) + \lambda \alpha^\top K \alpha + \mu^\top (K\alpha - u)$$

The dual function is given by $g(\mu) = \inf_{u,\alpha} L(u, \alpha, \mu)$

$$R^*(u) + \inf(\lambda \alpha^\top K \alpha + \mu^\top K \alpha)$$