Kernel Machines - Exam

Exercise 1: k psd, $\mathcal{H}$ its RKHS.

1.) PSD kernel: $k : \mathcal{X}^2 \to \mathbb{R}$ is psd if.

- "k_x" is symmetric: $k(x,y) = k(y,x)$, $\forall x,y \in \mathcal{X}$.
- $\forall n \in \mathbb{N}^*$, $\forall x_1, \dots, x_n \in \mathcal{X}$, $\forall \alpha_1, \dots, \alpha_n \in \mathbb{R}$:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j) \geq 0.$$

2 · Every psd kernel is a reproducing kernel.

we can define: $\mathcal{H} = \text{span}\{k(x,\cdot) \mid x \in \mathcal{X}\} = \{\sum_{i=1}^{n} a_i k_{x_i} \mid n \in \mathbb{N}^*, a \in \mathbb{R}^n, (x_i)_{1 \leq i \leq n} \in \mathcal{X}^n\}$.

we can define $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ as:

$$\langle \sum_{i=1}^{n} a_i k_{x_i}, \sum_{j=1}^{m} b_j k_{y_j} \rangle = \sum_{i,j} a_i b_j k(x_i, y_j).$$

we can verify $\mathcal{H}$ is an RKHS.

2) $G = \begin{bmatrix} (x^2+3)^5 & (xy+3)^5 \\ (xy+3)^5 & (y^2+3)^5 \end{bmatrix}$ is the Gram-matrix associated with the polynomial kernel $k : \mathbb{R}^2 \to \mathbb{R}$.

$$(x,y) \longmapsto k(x,y) = (\langle x,y \rangle + 3)^5$$

which is psd.

Hence G is positive semi-definite (symmetric) matrix.

3). a) using Cauchy-Schwarz:

$$|k'(x,y)| = \frac{|k(x,y)|}{\sqrt{k(x,x)}\sqrt{k(y,y)}} \leq \frac{\sqrt{k(x,x)} \cdot \sqrt{k(y,y)}}{\sqrt{k(x,x)}\sqrt{k(y,y)}} = 1$$

Hence $\sup_{x,y \in \mathbb{R}} |k'(x,y)| \leq 1$.

b) let $x \in \mathcal{X}$. using the reproducing property + Cauchy-Schwarz.

$$|g(x)| = |\langle g, k'_x \rangle_{\mathcal{H}}| \leq \|g\|_{\mathcal{H}} \underbrace{\|k'_x\|_{\mathcal{H}}}_{\leq 1} \leq \|g\|_{\mathcal{H}}.$$

so $\sup_{x \in \mathcal{X}} |f(x)| \leq \|g\|_{\mathcal{H}}$. $\Rightarrow$ f is bounded.

4) a) let $x,y \in \mathcal{X}$: $|k(x,y)| \leq \sqrt{k(x,x)}\sqrt{k(y,y)} = h(0)$

b) $\forall x_1, \dots, x_n \in \mathcal{X}$. $k_{x_i} \in \mathcal{H}$.

since $\mathcal{H}$ is a vector space (over $\mathbb{R}$). then $f = 1/n \sum_{i=1}^{n} k(x_i, \cdot) \in \mathcal{H}$.

Exercise 2:

1). Let X a random v. : $X \sim \mathbb{P}$.

The kernel k defines a unique feature map $\phi : \mathcal{X} \to \mathcal{H}$. s.t. : $\forall x \in \mathcal{X}$: $\phi(x) = k(x,\cdot) \in \mathcal{H}$.

Hence, we can define a mean element (by analogy to $\mathbb{E}(X)$ on $\mathcal{X}$):

$$\mu_{\mathbb{P}} = \mathbb{E}(\overset{\phi(x)}{k_x}) \quad \text{with } k_x := k(x,\cdot)$$

such that $\mu_{\mathbb{P}}(y) = \int_{\mathcal{X}} k(x,y)\, d\mathbb{P}(x)$, $\forall y \in \mathcal{X}$.

$$\Big(\mu_{\mathbb{P}}(y) = \langle \mathbb{E}(k_x), k_y \rangle_{\mathcal{H}} = \mathbb{E}\left(\langle k_x, k_y \rangle_{\mathcal{H}} \right) = \mathbb{E}\left(k(x,y) \right) \Big).$$

so $\mu_{\mathbb{P}}$ is the unique function in $\mathcal{H}$ verifying.

$\forall y \in \mathcal{X}$. $\boxed{\mu_{\mathbb{P}}(y) = \mathbb{E}(k(X,y))}$

2)
$$\mathbb{E}(f(X)) = \mathbb{E}(\langle f, k_X \rangle_{\mathcal{H}}) = \langle f, \mathbb{E}(k_X) \rangle_{\mathcal{H}} = \langle f, \mu_{\mathbb{P}} \rangle_{\mathcal{H}}.$$

3).
$$\max_{f \in \mathcal{H},\ \|f\|_{\mathcal{H}} \leq 1} \mathbb{E}(f(X)) \leq \max_{f \in \mathcal{H},\ \|f\|_{\mathcal{H}} \leq 1} \|f\|_{\mathcal{H}} \cdot \|\mu_{\mathbb{P}}\|_{\mathcal{H}} \leq \|\mu_{\mathbb{P}}\|_{\mathcal{H}}.$$

by taking $f = \frac{\mu_{\mathbb{P}}}{\|\mu_{\mathbb{P}}\|_{\mathcal{H}}}$: then f verifies. $\max \mathbb{E}($"

$$\mathbb{E}(f(x)) = \left\langle \frac{\mu_{\mathbb{P}}}{\|\mu_{\mathbb{P}}\|_{\mathcal{H}}}, \mu_{\mathbb{P}} \right\rangle_{\mathcal{H}} = \|\mu_{\mathbb{P}}\|_{\mathcal{H}}.$$

4) We are interested in minimizing

$$\min_{f \in \mathcal{H}} \sum_{i=1}^{n} \|k(x_i,\cdot) - f\|_{\mathcal{H}}^2 = \sum_{i=1}^{n} k(x_i,x_i) - 2f(x_i) + \|f\|_{\mathcal{H}}^2.$$

$$= \min_{f \in \mathcal{H}} \sum_{i=1}^{n} -2f(x_i) + \|f\|_{\mathcal{H}}^2 . = \min_{f \in \mathcal{H}} -2\sum_{i=1}^{n} f(x_i) + n\|f\|_{\mathcal{H}}^2.$$

→ Representer theorem: $f = \sum_{j=1}^{n} \alpha_j k_{x_j}$

$$\Leftrightarrow \min_{f \in \mathcal{H}} \sum_{i=1}^{n} \sum_{j=1}^{n} -2\alpha_j k(x_j, x_i) + n \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i,x_j).$$

$$\min_{\alpha \in \mathbb{R}^n} -2\alpha^T K \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + n\, \alpha^T K \alpha$$

we derive the expression w.r.t. α: $-2K\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = -2nK\alpha$

$$\Rightarrow \quad K\alpha = K \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix}.$$