CS-E4830 - Kernel Methods in Machine Learning D
Exercise 1

Alex Herrero Pons

Autumn Term Course 2020-2021

Question 1 (2 points): Recall from Lecture 1, the form for the polynomial kernel

$$K_1(x, y) = (\langle x, y\rangle + c)^m$$

where $c \geq 0$, m is a positive integer and $x, y \in \mathbb{R}^d$.

- Prove that $K_1(x, y)$ as defined above is a valid kernel.

Solution.
First of all we notice that $\langle x, y\rangle$ is a kernel by definition.
Then applying the Binomial Theorem $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$ we can express:

$$K_1(x, y) = (\langle x, y\rangle + c)^m = \sum_{k=0}^{m} \binom{m}{k} \langle x, y\rangle^{m-k} c^k$$

Now we know that the product of kernels is a kernel, therefore $\langle x, y\rangle^{m-k}$ is a kernel. We also know that $\binom{m}{k}$ and c^k are scalars, and a kernel multiplied by a scalar is a kernel too. Now what we have is a sum of kernels, and we also know that the conic sum of kernels is a kernel. Hence, $K_1(x, y)$ is a kernel.

$\square$

Question 2 (3 points) Recall from lecture 2, in the context of binary classification, the Parzen window classifier assigns a test instance x based on the distance to the centroids in the following way:

$$h(x) = \begin{cases} +1, & \text{if } ||\phi(x) - c_-||^2 > ||\phi(x) - c_+||^2 \\ -1, & \text{otherwise.} \end{cases}$$

where c_- and c_+ represent the centroids in the feature space of the negative and positive classes respectively. Show by deriving appropriate expressions for α_i and b, that the above decision function can be written in the following form $h(x) = \text{sgn}(\sum_{i=1}^{n} \alpha_i k(x, x_i) + b)$ such that $k(x, x_i) = \langle \phi(x), \phi(x_i) \rangle$. Here sgn(.) represents the sign function, and n is the total number of training samples.

Solution.

First of all we know that we can express $h(x)$ as:

$$h(x) = \text{sgn}\left(||\phi(x) - c_-||^2 - ||\phi(x) - c_+||^2\right)$$

We also know that $||x||^2 = \langle x, x \rangle$ therefore:

$$h(x) = \text{sgn}\left(\langle \phi(x) - c_-, \phi(x) - c_- \rangle - \langle \phi(x) - c_+, \phi(x) - c_+ \rangle\right)$$

By properties of the inner product we know that $\langle x - y, x - y \rangle = \langle x, x \rangle - 2\langle x, y \rangle + \langle y, y \rangle$. Hence:

$$\begin{aligned} h(x) = \text{sgn}\left(\langle \phi(x), \phi(x) \rangle - 2\langle \phi(x), c_- \rangle + \langle c_-, c_- \rangle - (\langle \phi(x), \phi(x) \rangle - 2\langle \phi(x), c_+ \rangle + \langle c_+, c_+ \rangle)\right) = \\ = \text{sgn}\left(-2\langle \phi(x), c_- \rangle + \langle c_-, c_- \rangle + 2\langle \phi(x), c_+ \rangle - \langle c_+, c_+ \rangle\right) = \\ = \text{sgn}\left(2\left(\langle \phi(x), c_+ \rangle - \langle \phi(x), c_- \rangle\right) + \langle c_-, c_- \rangle - \langle c_+, c_+ \rangle\right) = \\ = \text{sgn}\left(2\langle \phi(x), c_+ - c_- \rangle + \langle c_-, c_- \rangle - \langle c_+, c_+ \rangle\right) \end{aligned}$$

Then knowing that $c_- = \frac{1}{m_-} \sum_{i \in I^-} \phi(x_i)$ and $c_+ = \frac{1}{m_+} \sum_{i \in I^+} \phi(x_i)$:

$$\begin{aligned} h(x) = \text{sgn}\Bigg(2\left\langle \phi(x), \frac{1}{m_+} \sum_{i \in I^+} \phi(x_i) - \frac{1}{m_-} \sum_{i \in I^-} \phi(x_i) \right\rangle + \\ + \left\langle \frac{1}{m_-} \sum_{i \in I^-} \phi(x_i), \frac{1}{m_-} \sum_{i \in I^-} \phi(x_i) \right\rangle - \left\langle \frac{1}{m_+} \sum_{i \in I^+} \phi(x_i), \frac{1}{m_+} \sum_{i \in I^+} \phi(x_i) \right\rangle\Bigg) = \\ = \text{sgn}\Bigg(2\left\langle \phi(x), \sum_{i \in I^+} \frac{1}{m_+}\phi(x_i) + \sum_{i \in I^-} -\frac{1}{m_-}\phi(x_i) \right\rangle + \\ + \left\langle \frac{1}{m_-} \sum_{i \in I^-} \phi(x_i), \frac{1}{m_-} \sum_{i \in I^-} \phi(x_i) \right\rangle - \left\langle \frac{1}{m_+} \sum_{i \in I^+} \phi(x_i), \frac{1}{m_+} \sum_{i \in I^+} \phi(x_i) \right\rangle\Bigg) \end{aligned}$$

Now let $\alpha_i = \begin{cases} \frac{1}{m_+} & \text{if } y_i = +1 \\ -\frac{1}{m_-} & \text{if } y_i = -1 \end{cases}$

$$h(x) = \text{sgn}\Bigg(2\left\langle \phi(x), \sum_{i=0}^{m} \alpha_i \phi(x_i) \right\rangle +$$

$$+\left\langle \frac{1}{m_-}\sum_{i\in I^-}\phi(x_i), \frac{1}{m_-}\sum_{i\in I^-}\phi(x_i)\right\rangle - \left\langle \frac{1}{m_+}\sum_{i\in I^+}\phi(x_i), \frac{1}{m_+}\sum_{i\in I^+}\phi(x_i)\right\rangle\Bigg) =$$

$$= \operatorname{sgn}\Bigg(2\left\langle \phi(x), \sum_{i=0}^{m}\alpha_i\phi(x_i)\right\rangle +$$

$$+\frac{1}{m_-^2}\left\langle \sum_{i\in I^-}\phi(x_i), \sum_{i\in I^-}\phi(x_i)\right\rangle - \frac{1}{m_+^2}\left\langle \sum_{i\in I^+}\phi(x_i), \sum_{i\in I^+}\phi(x_i)\right\rangle\Bigg) =$$

$$= \operatorname{sgn}\left(2\sum_{i=0}^{m}\alpha_i k(x, x_i) + \frac{1}{m_-^2}\sum_{i,j\in I^-}k(x_i, x_j) - \frac{1}{m_+^2}\sum_{i,j\in I^+}k(x_i, x_j)\right)$$

We can divide everything inside the sgn function by 2 because it won't affect the sign of the result, therefore:

$$h(x) = \operatorname{sgn}\left(\sum_{i=0}^{m}\alpha_i k(x, x_i) + \frac{1}{2m_-^2}\sum_{i,j\in I^-}k(x_i, x_j) - \frac{1}{2m_+^2}\sum_{i,j\in I^+}k(x_i, x_j)\right)$$

Finally, let $b = \frac{1}{2m_-^2}\sum_{i,j\in I^-}k(x_i, x_j) - \frac{1}{2m_+^2}\sum_{i,j\in I^+}k(x_i, x_j)$:

$$h(x) = \operatorname{sgn}\left(\sum_{i=0}^{m}\alpha_i k(x, x_i) + b\right)$$

□

Question 3 (3 points) For $x, y \in \mathbb{R}$, check if $K_2(x, y) = cos(x + y)$ is a valid kernel function.

Solution. It's not a valid kernel. Prove with counter example.
We know that all kernels are positive definite functions*:

***Definition-Positive definite functions:**
A symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is positive definite if $\forall n \geq 1$, $\forall (a_1, \ldots, a_n) \in \mathbb{R}^n$, $\forall (x_1, \ldots, x_n) \in \mathcal{X}^n$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0$$

$K_2(x, y)$ is a symmetric function, but when $n = 1$ and i.e. $x = 120$ and $y = 0$:

$$a^2 K_2(x, y) = a^2 cos(x + y) = a^2 cos(120) < 0$$

We can see that $K_2(x, y) = cos(x + y)$ is not is a positive definite function, and therefore it's not a valid kernel.

$\square$

Question 4 (2 points) For $x, y \in \mathcal{X} = (-1, 1)$, prove that $K_3(x, y) = \frac{1}{1-xy}$ is a valid kernel.

Solution. It's a valid kernel. Prove with Mercer's Theorem*
First of all, having that $xy \in \mathcal{X} = (-1, 1)$ and applying Taylor's series we know that:

$$K_3(x, y) = \frac{1}{1 - xy} = 1 + xy + (xy)^2 + (xy)^3 + \cdots = \sum_{n=0}^{\infty} (xy)^n$$

*The Mercer's Theorem says that $K_3(x, y) = \frac{1}{1-xy}$ has to be

- Continuous,
- Symmetric,
- and Positive semi-definite

to be a valid kernel.
We can observe how the given function is continuous for all $x, y \in \mathcal{X} = (-1, 1)$. It's also symmetric because the product is symmetric. And also positive semi-definite because[1]:

$$\lim_{x \to \pm\infty} K_3(x, y) = \lim_{x \to \pm\infty} \frac{1}{1 - xy} = 0$$

$$\lim_{y \to \pm\infty} K_3(x, y) = \lim_{y \to \pm\infty} \frac{1}{1 - xy} = 0$$

Therefore,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j K_3(x_i, x_j) \geq 0$$

$$\forall n \geq 1, \forall (a_1, \ldots, a_n) \in \mathbb{R}^n, \forall (x_1, \ldots, x_n) \in \mathcal{X}^n$$

.

□

[1] I'm not sure if this proves positive semi-definiteness.