Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 6

A kernel function $k\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ must satisfy the *Mercer's condition*, which verifies that for any sequence of data points $x_1, \ldots, x_n \in \mathbb{R}^d$ and coefficients $c_1, \ldots, c_n \in \mathbb{R}$ the inequality

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

is satisfied. If it is the case, the kernel is called a *Mercer kernel*.

Conversely, the *representer theorem* states that if k is a Mercer kernel on $\mathbb{R}^d$, then there exists a Hilbert space (i.e., a finite or infinite dimensional $\mathbb{R}$-vector space with norm and scalar product) $\mathcal{F}$, the so-called feature space, and a continuous map $\varphi : \mathbb{R}^d \to \mathcal{F}$, such that

$$k(x, x') = \langle \, \varphi(x) \, , \, \varphi(x') \, \rangle_{\mathcal{F}} \quad \text{for all } x, x' \in \mathbb{R}^d.$$

Exercise 1: Mercer Kernels (3 × 20 P)

(a) *Show* that the following are Mercer kernels.

i. $k(x, x') = \langle x, x' \rangle$
ii. $k(x, x') = f(x) \cdot f(x')$ where $f\colon \mathbb{R}^d \to \mathbb{R}$ is an arbitrary continuous function

(b) Let k_1, k_2 be two Mercer kernels, for which we assume the existence of a finite-dimensional feature map associated to them. *Show* that the following are again Mercer kernels.

i. $k(x, x') = k_1(x, x') + k_2(x, x')$
ii. $k(x, x') = k_1(x, x') \cdot k_2(x, x')$

(c) *Show* using the results above that the polynomial kernel of degree d, where $k(x, x') = (\langle x, x' \rangle + \vartheta)^d$ and $\vartheta \in \mathbb{R}^+$, is a Mercer kernel.

Exercise 2: The Feature Map (4 × 10 P)

Consider the homogenous polynomial kernel k of degree 2 which is $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, where

$$k(x, y) = \langle x, y \rangle^2 = \Big(\sum_{i=1}^{2} x_i \, y_i \Big)^2.$$

(a) *Show* that $\mathcal{F} = \mathbb{R}^3$ and $\varphi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^2 \\ \sqrt{2} \, x_1 \, x_2 \\ x_2^2 \end{pmatrix}$ are possible choices for feature space and feature map.

(b) Consider the unit circle $C = \left\{ \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \; ; \; 0 \leq \theta < 2\pi \right\}$. *Show* that the image $\varphi(C)$ lies on a plane H in $\mathbb{R}^3$.

(c) Consider the plane $A = \left\{ \begin{pmatrix} t \\ s \end{pmatrix} \; ; \; t, s \in \mathbb{R} \right\}$. *Find* a point P in $\mathcal{F}$ which is not contained in $\varphi(A)$.

(d) *Find* a feature map associated to the kernel $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ with $k(x, y) = \langle x, y \rangle^2 = \Big(\sum_{i=1}^{d} x_i \, y_i \Big)^2$.

Sheet 7

1 a)

I $\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \langle x_i, x_j \rangle$

$\displaystyle = \langle \sum_{i=1}^{n} c_i x_i, \sum_{j=1}^{n} c_j x_j \rangle = \sum_{i=1}^{n} c_{i1} x_{i1} \sum_{j=1}^{n} c_{j1} x_{j1} + \cdots + \sum_{i=1}^{n} c_{id} x_{id} \sum_{j=1}^{n} c_{jd} x_{jd}$

$\displaystyle = \sum_{d=1}^{D} \sum_{i=1}^{n} c_{id} x_{id} \sum_{j=1}^{n} c_{jd} x_{jd}$

$\displaystyle = \sum_{d=1}^{D} \left(\sum_{m=1}^{n} c_{md} x_{md} \right)^2 \geq 0$

$\sum_{i=1}^{n}$ and $\sum_{j=1}^{n}$ are equivalent since $n = n$.

II $\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j f(x_i) \cdot f(x_j)$

$= c_{i1} f(x_{i1}) c_{j1} f(x_{j1}) + \cdots + c_{in} f(x_{in}) c_{jn} f(x_{jn})$

$= \left[c_{i1} f(x_{i1})\right]^2 + \cdots + \left[c_{in} f(x_{in})\right]^2 \geq 0$

since $i = j$ in all cases.

b) I $\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \left[k_1(x_i, x_j) + k_2(x_i, x_j)\right]$

$$= \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k_1(x_i, x_j)}_{\geq 0} + \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k_2(x_i, x_j)}_{\geq 0} \geq 0$$

II $\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k(x_i, x_j)$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k_1(x_i, x_j)\, k_2(x_i, x_j)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \langle \varphi_1(x_i), \varphi_1(x_j)\rangle \langle \varphi_2(x_i), \varphi_2(x_j)\rangle$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \sum_{m=1}^{d} \varphi_1(x_{im})\varphi_1(x_{jm}) \sum_{k=1}^{d} \varphi_2(x_{ik})\varphi_2(x_{jk})$$

$$= \sum_{m,k=1}^{d} \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \varphi_1(x_{im})\varphi_2(x_{ik})\varphi_1(x_{jm})\varphi_2(x_{jk})$$

$$= \sum_{m,k=1}^{d} \sum_{i=1}^{n} c_i \varphi_1(x_{im})\varphi_2(x_{ik}) \sum_{j=1}^{n} \varphi_1(x_{jm})\varphi_2(x_{jk})$$

$$= \sum_{m,k=1}^{d} \left(\sum_{i=1}^{n} c_i \varphi_1(x_{im})\varphi_2(x_{ik})\right)^2 \geq 0$$ since i = j in all cases.

c) We first show that $k(x, x') = \langle x, x' \rangle$ is a Mercer kernel:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \langle x_i, x_j \rangle$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \sum_{m=1}^{d} x_{im} x_{jm} = \sum_{m=1}^{d} \sum_{i=1}^{n} c_i x_{im} \sum_{j=1}^{n} c_j x_{jm}$$

$$= \sum_{m=1}^{d} \left(\sum_{i=1}^{n} c_i x_{im} \right)^2 \geq 0$$

We next show that $k(x, x') = \vartheta$ with $\vartheta \in \mathbb{R}^+$ is a Mercer kernel:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \vartheta = \sum_{i=1}^{n} c_i^2 \vartheta \geq 0$$

since $\vartheta > 0$ and $\sum_{i=1}^{n} c_i^2 \geq 0$.

Since $k_A(x, x') = \langle x, x' \rangle$ and $k_B = \vartheta$ are both Mercer kernels, $k_C(x, x') = k_A + k_B$ must be a Mercer kernel too according to exercise 1 b I.

If k_C is a Mercer kernel, k_C^d is a Mercer kernel too because the product of d Mercer kernels are again Mercer kernels according to exercise 1 b II.

Therefore,

$$k_D(x, x') = k_C^d = (k_A + k_B)^d = (\langle x, x' \rangle + \vartheta)^d$$

is a Mercer kernel for $\vartheta \in \mathbb{R}^+$.

(with $F = \mathbb{R}^3$)

$\boxed{2}$

a) The feature map $\varphi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^2 \\ \sqrt{2}\, x_1 x_2 \\ x_2^2 \end{pmatrix}$ is a possible choice for the kernel $k(\bar{x}, \bar{y}) = \langle \bar{x}, \bar{y} \rangle^2$

because

$$k(\bar{x}, \bar{y}) = \langle \varphi(\bar{x}), \varphi(\bar{y}) \rangle_F = \left\langle \begin{pmatrix} x_1^2 \\ \sqrt{2}\, x_1 x_2 \\ x_2^2 \end{pmatrix}, \begin{pmatrix} y_1^2 \\ \sqrt{2}\, y_1 y_2 \\ y_2^2 \end{pmatrix} \right\rangle$$

$$= x_1^2 y_1^2 + \sqrt{2}\, x_1 x_2 \sqrt{2}\, y_1 y_2 + x_2^2 y_2^2$$

$$= (x_1 y_1)^2 + 2\, x_1 y_1 x_2 y_2 + (x_2 y_2)^2$$

$$= (x_1 y_1 + x_2 y_2)^2 = \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle^2 = \langle \bar{x}, \bar{y} \rangle^2$$

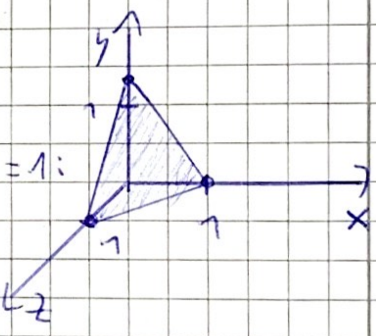
b)

$$\varphi(C) = \varphi\begin{pmatrix} \cos\vartheta \\ \sin\vartheta \end{pmatrix} = \begin{pmatrix} \cos^2\vartheta \\ \sqrt{2}\cos\vartheta \sin\vartheta \\ \sin^2\vartheta \end{pmatrix}$$

which lies on a plane in $\mathbb{R}^3$.

c) $\varphi(A) = \varphi\begin{pmatrix} t \\ s \end{pmatrix} = \begin{pmatrix} t^2 \\ \sqrt{2}\, ts \\ s^2 \end{pmatrix}$ if $t = s = 1$:

$\begin{pmatrix} 1^2 \\ \sqrt{2} \\ 1^2 \end{pmatrix} \longrightarrow$



The point $P = (-1, \sqrt{2}, -1)$ can not lie in this plane because the axes x and z can only contain positive numbers since $x \mapsto t^2$ and $z \mapsto s^2$ with $t, s \in \mathbb{R}$.

d) $\langle x, y \rangle^2 = \left(\sum_{i=1}^{d} x_i y_i \right)^2 = (x_1 y_1 + x_2 y_2 + \dots + x_d y_d)^2$

$= x_1^2 y_1^2 + \sqrt{2}\, x_1 x_2 \sqrt{2}\, y_1 y_2 + \dots + \sqrt{2}\, x_1 x_d \sqrt{2}\, y_1 y_d$

$+ x_2^2 y_2^2 + \sqrt{2}\, x_2 x_3 \sqrt{2}\, y_2 y_3 + \dots + \sqrt{2}\, x_2 x_d \sqrt{2}\, y_2 y_d$

$+ \dots$

$+ x_{d-1}^2 y_{d-1}^2 + \sqrt{2}\, x_{d-1} x_d \sqrt{2}\, y_{d-1} y_d$

$+ x_d^2 y_d^2$

This leads to

$$\varphi(x) = \begin{pmatrix} x_1^2 \\ \sqrt{2}\, x_1 x_2 \\ \sqrt{2}\, x_1 x_3 \\ \vdots \\ \sqrt{2}\, x_1 x_d \\ x_2^2 \\ \sqrt{2}\, x_2 x_3 \\ \sqrt{2}\, x_2 x_4 \\ \vdots \\ \sqrt{2}\, x_2 x_d \\ \vdots \\ x_{d-1}^2 \\ \sqrt{2}\, x_{d-1} x_d \\ x_d^2 \end{pmatrix}$$