Exercises for the course

Machine Learning 1

Winter semester 2024/25

Fachgebiet 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 5

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

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 $x_1, \dots, x_n \in \mathbb{R}^d$ and coefficients $c_1, \dots, c_n \in \mathbb{R}$ 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 $\times$ 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

i.

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

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \sum_{k=1}^{d} x_{i,k} \cdot x_{j,k}$$

$$= \sum_{k=1}^{d} \left(\sum_{i=1}^{n}\sum_{j=1}^{n} c_i x_{i,k} \cdot c_j x_{j,k} \right)$$

$$= \sum_{k=1}^{d} \Big(\underbrace{\sum_{i=1}^{n} c_i x_{i,k}}_{\text{they're same}} \cdot \underbrace{\sum_{j=1}^{n} c_j x_{j,k}}_{} \Big)$$

$$= \sum_{k=1}^{d} \left[\left(\sum_{i=1}^{n} c_i x_{i,k} \right)^2 \right] \geq 0$$

$$\sum_i \sum_j c_i c_j k(x_i, x_j) = \sum_i \sum_j c_i c_j \langle x_i, x_j \rangle$$

$$= \sum_i \sum_j \langle c_i \cdot x_i, c_j \cdot x_j \rangle$$

$$= \left\langle \sum_i c_i x_i, \sum_j c_j x_j \right\rangle$$

$$= \left\| \sum_i c_i x_i \right\|^2 \geq 0$$

Bilinearity:

$$\alpha \cdot \langle x, y \rangle = \langle \alpha \cdot x, y \rangle = \langle x, \alpha \cdot y \rangle$$

$$\langle x, y \rangle + \langle x, z \rangle = \langle x, y + z \rangle$$

$\varphi(x) = x$ $\qquad K = \Phi(X)\Phi(X)^T$

$$K_{ij} = \langle x_i, x_j \rangle \to c^T K c \geq 0$$
$$= \langle \Phi(x_i), \Phi(x_j) \rangle$$

ii.

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

$$= \underbrace{\sum_{i=1}^{n} c_i f(x_i)}_{\text{they're same}} \underbrace{\sum_{j=1}^{n} c_j f(x_j)}_{}$$

$$= \left(\sum_{i=1}^{n} c_i f(x_i) \right)^2 \geq 0$$

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

i.

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

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

because k_1, k_2 are Mercer Kernels

≥ 0

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

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

We assume φ_1, φ_2 project the data into the same dimension.
If they project into different dimension, we extend the lower-dimensional vector with zeros until the two projected dimensions are equal

Note that if there is a $\varphi: \mathbb{R}^d \to \mathbb{R}^a$ such that $k(x,x') = \langle \varphi(x), \varphi(x')\rangle$ for all $x, x' \in \mathbb{R}^d$
then there will also be a $\varphi': \mathbb{R}^d \to \mathbb{R}^{a+b}$ such that $k(x,x') = \langle \varphi'(x), \varphi'(x')\rangle$ for all $x, x' \in \mathbb{R}^d$.

$$\varphi'(x) = \begin{pmatrix} \varphi_1(x) \\ \vdots \\ \varphi_a(x) \\ 0 \\ \vdots \\ 0 \end{pmatrix} \left.\vphantom{\begin{matrix}0\\ \vdots \\ 0\end{matrix}}\right\} b \quad \text{, because } \langle \varphi'(x), \varphi'(x)\rangle = \langle \varphi(x), \varphi(x)\rangle.$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \left(\sum_{k=1}^{d} \varphi_{1,k}(x_i)\varphi_{1,k}(x_j)\right)\left(\sum_{l=1}^{d} \varphi_{2,l}(x_i)\varphi_{2,l}(x_j)\right)$$

$$= \sum_{k=1}^{d}\sum_{l=1}^{d} \left(\sum_{i=1}^{n} c_i \varphi_{1,k}(x_i)\varphi_{2,l}(x_i) \cdot \sum_{j=1}^{n} c_j \varphi_{1,k}(x_j)\varphi_{2,l}(x_j)\right)$$

$$= \sum_{k=1}^{d}\sum_{l=1}^{d} \left[\left(\sum_{i=1}^{n} c_i \varphi_{1,k}(x_i)\varphi_{2,l}(x_i)\right)^2\right] \geq 0$$

(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.

First, we prove that $k(x,x') = \vartheta$, $\vartheta \in \mathbb{R}^+$ is a Mercer Kernel

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

$$= \vartheta \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j$$

$$= \vartheta \cdot \sum_{i=1}^{n} c_i \cdot \sum_{j=1}^{n} c_j$$

$$= \vartheta \cdot \left(\sum_{i=1}^{n} c_i\right)^2 \geq 0$$

$\Rightarrow k(x,x') = \vartheta$, $\vartheta \in \mathbb{R}^+$ is a Mercer Kernel

Now we give a proof by induction

- Base Case: We show that $k(x,x') = (\langle x,x'\rangle + \vartheta)^d$ is a Mercer Kernel for $d = 1$
  i.e. $k(x,x') = \langle x,x'\rangle + \vartheta$

$k(x,x') = \langle x,x'\rangle + \vartheta$ is clearly true because $\langle x,x'\rangle$ and ϑ are Mercer kernel. And the Sum of two Mercer kernel is also a Mercer kernel. (from b.i)

- induction step: We show that for every $d \geq 1$. if $k(x,x') = (\langle x,x'\rangle + \vartheta)^d$ is Mercer Kernel,
  then $k(x,x') = (\langle x,x'\rangle + \vartheta)^{d+1}$ also is a Mercer Kernel

Assume the induction hypothesis that $k(x,x') = (\langle x,x'\rangle + \vartheta)^d$ is Mercer Kernel

$$k(x,x') = (\langle x,x'\rangle+\vartheta)^{d+1} = \underbrace{(\langle x,x'\rangle+\vartheta)^d}_{\text{Mercel Kernel}} \cdot \underbrace{(\langle x,x'\rangle+\vartheta)}_{\text{Mercel kernel}}$$

$$\Rightarrow k(x,x') = (\langle x,x'\rangle+\vartheta)^{d+1} \text{ is Mercel kernel}$$

- Conclusion: Since both the base case and the induction step have been proved as true,

$$k(x,x') = (\langle x,x'\rangle+\vartheta)^d \text{ is Mercer kernel for } \vartheta \in \mathbb{R}^+ \text{ and } d \in \mathbb{N}^+$$

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.

$$\begin{aligned}
k(x,y) = \langle \varphi(x), \varphi(y)\rangle &= \left\langle \varphi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \varphi\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right\rangle \\
&= \left\langle \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix}, \begin{pmatrix} y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \end{pmatrix}\right\rangle \\
&= x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 \\
&= (x_1y_1 + x_2y_2)^2 \\
&= \Big(\sum_{i=1}^{2} x_iy_i\Big)^2 = \langle x,y\rangle^2
\end{aligned}$$

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

$$\varphi(C) = \varphi\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} \cos^2\theta \\ \sqrt{2}\cos\theta\sin\theta \\ \sin^2\theta \end{pmatrix} \quad \text{for } 0 \le \theta < 2\pi$$

$$= \left\{ \begin{pmatrix} \cos^2\theta \\ \sqrt{2}\cos\theta\sin\theta \\ 1-\cos^2\theta \end{pmatrix} ; 0 \le \theta \le 2\pi \right\}$$

$$= \left\{ \begin{pmatrix} t \\ s \\ 1-t \end{pmatrix} ; s,t \in \mathbb{R} \right\}$$

$$= \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

The following three points are on the plane

$$\varphi\left(\begin{pmatrix} \cos 0 \\ \sin 0 \end{pmatrix}\right) = \varphi\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\varphi\left(\begin{pmatrix} \cos \frac{1}{2}\pi \\ \sin \frac{1}{2}\pi \end{pmatrix}\right) = \varphi\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\varphi\left(\begin{pmatrix} \cos(\frac{1}{4}\pi) \\ \sin(\frac{1}{4}\pi) \end{pmatrix}\right) = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} = \begin{pmatrix} (\sqrt{2}/2)^2 \\ \sqrt{2}\cdot(\sqrt{2}/2)\cdot(\sqrt{2}/2) \\ (\sqrt{2}/2)^2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ \sqrt{2}/2 \\ 1/2 \end{pmatrix}$$

The plane can be written in the form

$$H: x = \begin{pmatrix}1\\0\\0\end{pmatrix} + r \cdot \left[\begin{pmatrix}0\\0\\1\end{pmatrix} - \begin{pmatrix}1\\0\\0\end{pmatrix}\right] + s \cdot \left[\begin{pmatrix}1/2\\\sqrt{2}/2\\1/2\end{pmatrix} - \begin{pmatrix}1\\0\\0\end{pmatrix}\right]$$

$$= \begin{pmatrix}1\\0\\0\end{pmatrix} + r \cdot \begin{pmatrix}-1\\0\\1\end{pmatrix} + s \cdot \begin{pmatrix}-1/2\\\sqrt{2}/2\\1/2\end{pmatrix}$$

We can rewrite this into the normal-vector representation

$$H: \left(\begin{pmatrix}-1\\0\\1\end{pmatrix} \times \begin{pmatrix}-1/2\\\sqrt{2}/2\\1/2\end{pmatrix}\right)^T \cdot \left(x - \begin{pmatrix}1\\0\\0\end{pmatrix}\right) = \begin{pmatrix}-\sqrt{2}/2\\0\\-\sqrt{2}/2\end{pmatrix}^T \cdot \left(x - \begin{pmatrix}1\\0\\0\end{pmatrix}\right)$$

(the cross product of two vector $\vec{u}$ and $\vec{v}$ results in a vector which is othogonal to $\vec{u}$ and $\vec{v}$)

The formula returns 0 for all points on the plane.

$$\begin{pmatrix}-\sqrt{2}/2\\0\\-\sqrt{2}/2\end{pmatrix} \cdot \left(\varphi(x) - \begin{pmatrix}1\\0\\0\end{pmatrix}\right) = \begin{pmatrix}-\sqrt{2}/2\\0\\-\sqrt{2}/2\end{pmatrix}^T \cdot \left(\begin{pmatrix}\cos(\theta)^2\\\sqrt{2}\cos(\theta)\sin(\theta)\\\sin(\theta)^2\end{pmatrix} - \begin{pmatrix}1\\0\\0\end{pmatrix}\right)$$

$$= \begin{pmatrix}-\sqrt{2}/2\\0\\-\sqrt{2}/2\end{pmatrix}^T \cdot \begin{pmatrix}\cos(\theta)^2 - 1\\\sqrt{2}\cos(\theta)\sin(\theta)\\\sin(\theta)^2\end{pmatrix}$$

$$= -\sqrt{2}/2 \cdot (\cos(\theta)^2 - 1) - \sqrt{2}/2 \sin(\theta)^2$$

$$= -\sqrt{2}/2 \cdot (\underbrace{\cos(\theta)^2 + \sin(\theta)^2}_{1} - 1)$$

$$= -\sqrt{2}/2 \cdot (1-1)$$

$$= 0$$

Therefore the image $\varphi(C)$ lies on the 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 $\varphi(A)$.

$$\varphi(A) = \varphi\begin{pmatrix} t \\ s \end{pmatrix} = \begin{pmatrix} t^2 \\ \sqrt{2}\, t s \\ s^2 \end{pmatrix}$$

if we choose $t = 1$ and $s = 1$, clearly we have $\varphi\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix}$

Then alle the Point $\begin{pmatrix} 1 \\ k \\ 1 \end{pmatrix}$ with $k \in \mathbb{R} \setminus \{\sqrt{2}\}$ are not contained in $\varphi(A)$. e.g. $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

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

−1

$$k = \langle x, y \rangle^2 = (x^T \cdot y)^2 = \left[(x_1\ x_2\ \cdots\ x_d) \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} \right]^2$$

$$= (x_1 y_1 + x_2 y_2 + \dots + x_d y_d)^2$$

$$= (x_1 y_1)^2 + \dots + (x_d y_d)^2 + 2 x_1 y_1 x_2 y_2 + \dots + 2 x_d y_d x_{d-1} y_{d-1}$$

$$= \left\langle \begin{pmatrix} x_1^2 \\ \sqrt{2}\, x_1 x_2 \\ \sqrt{2}\, x_1 x_3 \\ \vdots \\ \sqrt{2}\, y_d y_{d-1} \\ x_d^2 \end{pmatrix}, \begin{pmatrix} y_1^2 \\ \sqrt{2}\, y_1 y_2 \\ \sqrt{2}\, y_1 y_3 \\ \vdots \\ \sqrt{2}\, y_d y_{d-1} \\ y_d^2 \end{pmatrix} \right\rangle$$

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

$\phi(x) = [(x_i^2)_i, (\sqrt{2}\, x_i x_j)_{i<j}] \in \mathbb{R}^{d \cdot (d+1)/2}$ →

$\phi(x) = [(x_i x_j)_{ij}] \in \mathbb{R}^{d^2}$ → $\langle \Phi(x), \Phi(y) \rangle \sim \mathcal{O}(d^2)$

→ $\langle x, y \rangle^2 \sim \mathcal{O}(d)$