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

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

i. $\langle x, x' \rangle = \sum_{k=1}^{d} x_k x'_k$ with d is the dimension of x.

$$\underline{\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \langle x_i, x_j \rangle} = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \sum_{k=1}^{d} x_{ik} x_{jk}$$

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

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

with $\sum_{i=1}^{n} c_i x_{ik} = \sum_{j=1}^{n} c_j x_{jk}$,

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

ii. similar like above,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j f(x_i) \cdot f(x_j) = \left(\sum_{i=1}^{n} c_i f(x_i)\right)\left(\sum_{j=1}^{n} c_j f(x_j)\right)$$

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

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

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

with k_1, k_2 are already Mercer kernels,

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

so $\underline{\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \left(k_1(x_i, x_j) + k_2(x_i, x_j) \right) \geq 0}$ ✓

ii. $\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k_1(x_i, x_j) k_2(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \langle \varphi_1(x_i), \varphi_1(x_j) \rangle \langle \varphi_2(x_i), \varphi_2(x_j) \rangle$

with $\langle x, x' \rangle = \sum_{k=1}^{d} x_k, x_k'$,

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

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

since $\sum_{i=1}^{n} c_i \varphi_{1k_i}(x_i) \varphi_{2l_i}(x_i) = \sum_{j=1}^{n} c_j \varphi_{1k_j}(x_j) \varphi_{2l_j}(x_j)$

$$\underline{\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k_1(x_i, x_j) k_2(x_i, x_j) = \sum_{k=1}^{d}\sum_{l=1}^{d} \left(\sum_{i=1}^{n} c_i \varphi_{1k_i}(x_i) \varphi_{2l_i}(x_i) \right)^2 \geq 0}$$ ✓

(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.

$$k(x,x') = (\langle x,x' \rangle + \vartheta)^d$$

$$= \underbrace{(\langle x,x' \rangle + \sqrt{\vartheta} \cdot \sqrt{\vartheta})}_{1} \cdot \underbrace{(\langle x,x' \rangle + \sqrt{\vartheta} \cdot \sqrt{\vartheta})}_{2} \cdot \dots \cdot \underbrace{(\langle x,x' \rangle + \sqrt{\vartheta} \cdot \sqrt{\vartheta})}_{d}$$

$\vartheta \in \mathbb{R}^+$, then $\sqrt{\vartheta}$ is a continuous function.

from (a) (b) we already know that:

$\langle x,x' \rangle$, $f(x) \cdot f(x')$, $k_1(x,x') + k_2(x,x')$ and $k_1(x,x') \cdot k_2(x,x')$

are all Mercer kernels.

so $k(x,x') = (\langle x,x' \rangle + \vartheta)^d$ is also 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 $C = \left\{ \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \; ; \; 0 \le \theta < 2\pi \right\}$. *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) $k(x,y) = \langle \varphi(x) \cdot \varphi(y) \rangle$

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

$$= (x_1 y_1 + x_2 y_2)^2 = \Big(\sum_{i=1}^{2} x_i y_i \Big)^2$$

(b) $\varphi(c) = \begin{pmatrix} \cos^2\theta \\ \sqrt{2} \cos\theta \sin\theta \\ \sin^2\theta \end{pmatrix}$, with $\cos^2\theta + \sin^2\theta = 1$, $\cos\theta \cdot \sin\theta = \frac{1}{2} \sin(2\theta)$

so $\cos^2\theta = 1 - \sin^2\theta$, $\sin^2\theta$ and $\frac{\sqrt{2}}{2} \sin(2\theta) \in \mathbb{R}$,

and $\varphi(c)$ lies on a plane in $\mathbb{R}^3$.

c) $\varphi(A) = \begin{pmatrix} t^2 \\ \sqrt{2}\, ts \\ s^2 \end{pmatrix}$. let's take a point $P = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}$

when $t = s = 1$, $\varphi(A) = \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix}$

then $P \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}$ is not contained in $\varphi(A)$

(d) $k(x,y) = \left(\sum_{i=1}^{d} x_i y_i\right)^2 = (x_1y_1 + x_2y_2 + \dots + x_dy_d)$

$$= \sum x_i^2 y_i^2 + 2 \sum_i \sum_j x_i x_j \, y_i y_j$$

$$= \begin{pmatrix} x_1^2 \\ \vdots \\ x_d^2 \\ \sqrt{2}\, x_1 x_2 \\ \vdots \\ \sqrt{2}\, x_{d-1} x_d \end{pmatrix} \begin{pmatrix} y_1^2 \\ \vdots \\ y_d^2 \\ \sqrt{2}\, y_1 y_2 \\ \vdots \\ \sqrt{2}\, y_{d-1} y_d \end{pmatrix}$$

$$= \varphi(x)\,\varphi(y)$$

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