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 8

A kernel function $k\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ generalizes the linear scalar product between two vectors and allows to treat non-linearities in data by linear means. In order for that to work, the kernel function must fulfill certain properties which are shared by the scalar product. In particular, it 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

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

I.e., the kernel k implicitly allows to calculate scalar products in the feature space $\mathcal{F}$ without need for an algorithm to calculate φ. This scalar product form can also be useful as an alternate representation (φ-representation) of Mercer kernels, in particular, when showing whether a particular composition of Mercer kernels is or isn't a Mercer kernel itself.

Exercise 1: Kernology (4 × 15 P)

(a) *Show* that

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

are Mercer kernels.

(b) Let k_1, k_2 be two Mercer kernels. *Show* that

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

are again Mercer kernels.

(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.

(d) *Show* using the results above that the Gaussian kernel of width σ, where $k(x, x') = \exp\big(-\frac{\|x-x'\|^2}{2\sigma^2}\big)$ 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

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

(a) *Show* that $\mathcal{F} = \mathbb{R}^3$ and $\varphi : \mathbb{R}^2 \to \mathbb{R}^3$ where

$$\varphi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^2 \\ \sqrt{2} \, x_1 \, x_2 \\ x_2^2 \end{pmatrix}$$

are possible choices for feature space and feature map.

(b) *Determine* (i.e., give an explicit description of) the images of

i. The unit circle $C = \{ \binom{\cos\theta}{\sin\theta} \; ; \; 0 \leq \theta < 2\pi \}$
ii. The plane $A = \{ \binom{t}{s} \; ; \; t, s \in \mathbb{R} \}$

under the feature map φ in the feature space $\mathcal{F}$.

(c) The image $\varphi(C)$ lies on a plane H in $\mathbb{R}^3$. *Characterize* that plane.

(d) *Find* a point P in $\mathcal{F}$ which is not contained in $\varphi(A)$.