G. Montavon
Institute of Computer Science
Department of Mathematics and Computer Science
Freie Universität Berlin

Exercise Sheet 10 (theory part)

Exercise 1: Mercer Kernels (5 + 5 + 5 + 5 + 5 P)

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*.

(a) *Show* that $k(x, x') = \langle x, x' \rangle$ is a Mercer kernel.

$$\begin{aligned}
\textstyle\sum_{ij} c_i c_j k(x_i, x_j) &= \textstyle\sum_{ij} c_i c_j \langle x_i, x_j \rangle \\
&= \textstyle\sum_{ij} c_i c_j \sum_k x_{ik} x_{jk} \\
&= \textstyle\sum_k (\sum_i c_i x_{ik})(\sum_j c_j x_{jk}) \\
&= \textstyle\sum_k (\sum_i c_i x_{ik})^2 \geq 0
\end{aligned}$$

(b) *Show* that $k(x, x') = f(x) \cdot f(x')$ where $f\colon \mathbb{R}^d \to \mathbb{R}$ is an arbitrary continuous function, is a Mercer kernel.

$$\begin{aligned}
\textstyle\sum_{ij} c_i c_j k(x_i, x_j) &= \textstyle\sum_{ij} c_i c_j f(x_i) f(x_j) \\
&= (\textstyle\sum_i c_i f(x_i))(\sum_j c_j f(x_j)) \\
&= (\textstyle\sum_i c_i f(x_i))^2 \geq 0
\end{aligned}$$

(c) 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 $k(x, x') = k_1(x, x') + k_2(x, x')$ is a Mercer kernel.

$$\begin{aligned}
\textstyle\sum_{ij} c_i c_j k(x_i, x_j) &= \textstyle\sum_{ij} c_i c_j (k_1(x_i, x_j) + k_2(x_i, x_j)) \\
&= \underbrace{\textstyle\sum_{ij} c_i c_j k_1(x_i, x_j)}_{\geq 0} + \underbrace{\textstyle\sum_{ij} c_i c_j k_2(x_i, x_j)}_{\geq 0} \geq 0
\end{aligned}$$

where in the last line, we have exploited the fact that k_1 and k_2 are Mercer kernels.

(d) *Show* that $k(x, x') = k_1(x, x') \cdot k_2(x, x')$ is a Mercer kernel.

$$\begin{aligned}
\textstyle\sum_{ij} c_i c_j k(x_i, x_j) &= \textstyle\sum_{ij} c_i c_j k_1(x_i, x_j) k_2(x_i, x_j) \\
&= \textstyle\sum_{ij} c_i c_j \langle \phi_1(x_i), \phi_1(x_j) \rangle \langle \phi_2(x_i), \phi_2(x_j) \rangle \\
&= \textstyle\sum_{ij} c_i c_j \sum_m \phi_{1m}(x_i) \phi_{1m}(x_j) \sum_n \phi_{2n}(x_i) \phi_{2n}(x_j) \\
&= \textstyle\sum_m \sum_n \sum_i c_i \phi_{1m}(x_i) \phi_{2n}(x_i) \sum_j c_j \phi_{1m}(x_j) \phi_{2n}(x_j) \\
&= \textstyle\sum_m \sum_n (\sum_i c_i \phi_{1m}(x_i) \phi_{2n}(x_i))^2 \geq 0
\end{aligned}$$

(e) *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.

$$k(x,x') = \underbrace{\underbrace{\underbrace{(\underbrace{\langle x,x'\rangle}_{\text{kernel}} + \underbrace{\sqrt{\vartheta}\sqrt{\vartheta}}_{\text{kernel}})}_{\text{kernel}} \cdot (\langle x,x'\rangle + \sqrt{\vartheta}\sqrt{\vartheta})}_{\text{kernel}} \cdot \ldots \cdot (\langle x,x'\rangle + \sqrt{\vartheta}\sqrt{\vartheta})}_{\text{kernel}}$$

Exercise 2: The Feature Map (5 + 5 + 5 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.

$$\begin{aligned}
\langle x,y\rangle^2 &= (x_1y_1 + x_2y_2)^2 \\
&= x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2 \\
&= x_1^2y_1^2 + \sqrt{2}x_1x_2\sqrt{2}y_1y_2 + x_2^2y_2^2 \\
&= \underbrace{\begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix}}_{\varphi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} \cdot \underbrace{\begin{pmatrix} y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \end{pmatrix}}_{\varphi\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}}
\end{aligned}$$

(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 $\mathbb{R}^3$.

$$\begin{aligned}
\phi(C) &= \left\{\begin{pmatrix} \cos^2(\theta) \\ \sqrt{2}\cos(\theta)\sin(\theta) \\ \sin^2(\theta) \end{pmatrix} \;;\; 0 \le \theta < 2\pi\right\} \\
&= \left\{\begin{pmatrix} \cos^2(\theta) \\ \sqrt{2}\cos(\theta)\sin(\theta) \\ 1 - \cos^2(\theta) \end{pmatrix} \;;\; 0 \le \theta < 2\pi\right\} \\
&\subseteq \left\{\begin{pmatrix} t \\ s \\ 1-t \end{pmatrix} \;;\; t,s \in \mathbb{R}\right\}
\end{aligned}$$

(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 $\varphi(A)$.

$$\phi(A) = \left\{\begin{pmatrix} t^2 \\ \sqrt{2}ts \\ s^2 \end{pmatrix} ; t,s \in \mathbb{R}\right\} \not\ni \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$$

Exercise 3: Weighted Degree Kernels (10 + 5 + 5 P)

The weighted degree kernel has been proposed to represent DNA sequences ($\mathcal{A} = \{\mathrm{G}, \mathrm{A}, \mathrm{T}, \mathrm{C}\}$), and is defined for pairs of sequences of length L as:

$$k(x, z) = \sum_{m=1}^{M} \beta_m \sum_{l=1}^{L+1-m} I(u_{l,m}(x) = u_{l,m}(z)).$$

where $\beta_1, \ldots, \beta_M \geq 0$ are weighting coefficients, and where $u_{l,m}(x)$ is a substring of x which starts at position l and of length m. The function $I(.)$ is an indicator function which returns 1 if the input argument is true and 0 otherwise.

```
       x  AAACAAATAAGTAACTAATCTTTTAGGAAGAACGTTTCAACCATTTTGAG
#1-mers   .|.|.|||.|..||.|.|..|||.||...|....|...|||.......|..
#2-mers   .....||.....|.......||..|.............||..........
#3-mers   .....|..............|...|.............|...........
      x'  TACCTAATTATGAAATTAAATTTCAGTGTGCTGATGGAAACGGAGAAGTC
```

(a) *Show* that k is a positive semi-definite kernel. That is, show that

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

for all inputs $x_1, \ldots, x_N$ and choice of real numbers $c_1, \ldots, c_N$.

$$\begin{aligned}
&\sum_i \sum_j c_i c_j \sum_m \beta_m \sum_l I(u_{l,m}(x_i) = u_{l,m}(x_j)) \\
&\quad = \sum_i \sum_j c_i c_j \sum_m \beta_m \sum_l \sum_{w \in \mathcal{W}} I(u_{l,m}(x_i) = w) \cdot I(u_{l,m}(x_j) = w) \\
&\quad = \sum_m \sum_l \sum_{w \in \mathcal{W}} \Big(\sqrt{\beta_m} \sum_i c_i I(u_{l,m}(x_i) = w)\Big)\Big(\sqrt{\beta_m} \sum_j c_j I(u_{l,m}(x_j) = w)\Big) \\
&\quad = \underbrace{\sum_m \sum_l \sum_{w \in \mathcal{W}} \underbrace{\Big(\sqrt{\beta_m} \sum_i c_i I(u_{l,m}(x_i) = w)\Big)^2}_{\geq 0}}_{\geq 0}
\end{aligned}$$

(b) *Give* a feature map associated to this kernel for the special case $M = 1$.

Observe that

$$k(x, x') = \sum_{m=1}^{M} \beta_m \sum_{l=1}^{L+1-m} \sum_{w \in \mathcal{W}} I(u_{l,m}(x) = w) \cdot I(u_{l,m}(x') = w) \tag{1}$$

for $M = 1$, Eq. (1) reduces to

$$k(x, x') = \beta_1 \sum_{l=1}^{L} \sum_{w \in \{A,C,T,G\}} I(u_{l,1}(x) = w) \cdot I(u_{l,1}(x') = w)$$

From this expression, we can extract the feature map

$$\phi(x) = \Big(\big(\sqrt{\beta_1} I(u_{l,1}(x) = w)\big)_{w \in \{A,C,T,G\}}\Big)_{l=1}^{L}$$

which is $L \times 4$ dimensional.

(c) *Give* a feature map associated to this kernel for the special case $M = 2$ with $\beta_1 = 0$ and $\beta_2 = 1$.

for $M = 2$, Eq. (1) reduces to

$$k(x, x') = \sum_{l=1}^{L-1} \sum_{w \in \{A,C,T,G\}^2} I(u_{l,2}(x) = w) \cdot I(u_{l,2}(x') = w)$$

From this expression, we can extract the feature map

$$\phi(x) = \Big(\big(I(u_{l,2}(x) = w) \big)_{w \in \{A,C,T,G\}^2} \Big)_{l=1}^{L-1}$$

which is $(L-1) \times 16$ dimensional.