ML1 (WS 2016/17)
Exercise Sheet 1

Group SWVTI

(Gaurav Vashish, Nhat Tran, René Saitenmacher, Sebastian Walther, Matthias Ivantsits)

Exercise 1:

(a)

(i) Let $k(x, x') := a$ for some $a \in \mathbb{R}^+$. We verify that Mercer's condition holds:

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

(ii) Let $k(x, x') := \langle x, x' \rangle$. We verify that Mercer's condition holds:

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k(x_i, x_j) \\
=&\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \langle x_i, x_j \rangle \\
=&\sum_{i=1}^{n} c_i \langle x_i, \sum_{j=1}^{n} c_j x_j \rangle \\
=&\langle \sum_{i=1}^{n} c_i x_i, \sum_{j=1}^{n} c_j x_j \rangle \\
=&|| \sum_{i=1}^{n} c_i x_i ||^2 \geq 0
\end{aligned}$$

(iii) Let $k(x, x') := f(x)f(x')$ for some arbitrary, continuous function $f : \mathbb{R}^d \to \mathbb{R}$. We verify that Mercer's condition holds:

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

(b)

Let $k_1(x, x')$ and $k_2(x, x')$ be Mercer kernels.
(i) Let $k(x, x') := k_1(x, x') + k_2(x, x')$. We verify that Mercer's condition holds:

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

(ii) Let $k(x,x') := k_1(x,x')k_2(x,x')$, and let $\phi_1 : \mathbb{R}^d \to \mathbb{R}^{d_1}$ and $\phi_2 : \mathbb{R}^d \to \mathbb{R}^{d_2}$ be the feature space mappings for k_1 and k_2, respectively. Then,

$$\begin{aligned}
k(x,x') &= k_1(x,x')k_2(x,x') \\
&= \langle \phi_1(x), \phi_1(x') \rangle_{\mathbb{R}^{d_1}} \langle \phi_2(x), \phi_2(x') \rangle_{\mathbb{R}^{d_2}} \\
&= \left(\sum_{i=1}^{d_1} (\phi_1(x))_i \, (\phi_1(x'))_i \right) \left(\sum_{j=1}^{d_2} (\phi_2(x))_j \, (\phi_2(x'))_j \right) \\
&= \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} (\phi_1(x))_i \, (\phi_1(x))_j \, (\phi_1(x'))_i \, (\phi_1(x'))_j \\
&= \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} (\phi(x))_{\pi(i,j)} \, (\phi(x'))_{\pi(i,j)} \\
&= \sum_{k=1}^{d_3 := d_1 d_2} (\phi(x))_k \, (\phi(x'))_k \\
&= \langle \phi(x), \phi(x') \rangle_{\mathbb{R}^{d_3}}
\end{aligned}$$

with $\phi(x) := \phi_1(x)\phi_2(x)$ and some bijective $\pi : [d_1] \times [d_2] \to [d_3]$. As in $1a_{ii}$) we can use the bilinearity of the scalar product $\langle \phi(x), \phi(x') \rangle_{\mathbb{R}^{d_3}}$ to conclude that $k(x,x')$ fulfils Mercer's condition.

c)

According to $1b_i$) the sum of two Mercer kernels is again a Mercer kernel. $\vartheta \in \mathbb{R}^+$ and $\langle x, x' \rangle$ are Mercer kernels according to $1a_i$) and $1a_{ii}$). Thus, $\langle x, x' \rangle + \vartheta$ is a Mercer kernel. $k(x,x') = (\langle x, x' \rangle + \vartheta)^d, \vartheta \in \mathbb{R}^+$ can now be decomposed as:

$$\begin{aligned}
k(x,x') &= \prod_{i=1}^{d} (\langle x, x' \rangle + \vartheta) \\
&= \underbrace{(\langle x, x' \rangle + \vartheta) \cdot (\langle x, x' \rangle + \vartheta) \ldots (\langle x, x' \rangle + \vartheta)}_{d\text{-times}} \qquad \vartheta \in \mathbb{R}^+
\end{aligned}$$

According to $1b_{ii}$ the product of Mercer kernels is again a Mercer kernel. Therefore $k(x,x') = (\langle x, x' \rangle + \vartheta)^d, \vartheta \in \mathbb{R}^+$ is a Mercer kernel.

d)

$k(x,x')$ can be decomposed as follows:

$$\begin{aligned}
k(x,x') &= exp(-\tfrac{||x-x'||^2}{2\sigma^2}) \\
&= exp(\tfrac{-||x||^2}{2\sigma^2}) exp(\tfrac{-||x'||^2}{2\sigma^2}) exp(\tfrac{\langle x, x' \rangle}{\sigma^2})
\end{aligned}$$

Define $f(\hat{x}) = exp(\frac{-||\hat{x}||^2}{2\sigma^2})$. This leads to:

$$k(x,x') = f(x)f(x')exp(\tfrac{1}{\sigma^2} \langle x, x' \rangle)$$

$f(x)f(x')$ is a Mercer kernel, see $1a_{iii}$). We now show that $exp(\frac{1}{\sigma^2}\langle x, x'\rangle)$ is a Mercer kernel. $\frac{1}{\sigma^2}\langle x, x'\rangle$ is a valid Mercer kernel according to $1a_i$) (since $\sigma \geq 0$), $1a_{ii}$) and $1b_{ii}$). Since the exponential function is the sum of infinite polynomials (Taylor Series Expansion), it is:

$$exp(k(x, x')) = \sum_{i=0}^{\infty} \frac{1}{i!} k(x, x')^i$$

Using $1a_i$) as well as the sum and the product property of Mercer kernels, $exp(k(x, x'))$ is again a Mercer kernel. Therefore $k(x, x') = f(x)f(x')exp(\frac{1}{\sigma^2}\langle x, x'\rangle) = exp(-\frac{||x-x'||^2}{2\sigma^2})$ is also a valid Mercer kernel.

Exercise 2:

(a)

$$\begin{aligned}
k(x, x') &= \left(\sum_{i=1}^{d=2} x_i y_i\right)^2 \\
&= (x_1 y_1 + x_2 y_2)^2 \\
&= x_1^2 y_1^2 + 2x_1 y_1 x_2 y_2 + x_2^2 y_2^2 \\
&= \begin{pmatrix} x_1^2 \\ \sqrt{2} x_1 x_2 \\ x_2^2 \end{pmatrix}^T \begin{pmatrix} y_1^2 \\ \sqrt{2} y_1 y_2 \\ y_2^2 \end{pmatrix} \\
&= \langle \varphi(x), \varphi(y) \rangle_{\mathcal{F}}
\end{aligned}$$

(b)

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

(c)

Using the identity $\sin^2(\theta) + \cos^2(\theta) = 1$, we see that $\phi(C)$ lies on the plane anchored at some point $r = \begin{pmatrix} x \\ y \\ 1-x \end{pmatrix}$ and with normal vector $n = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ which can be described by the following linear equation:

$$x_1 + x_3 - 1 = 0$$

(d)

$\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ is not contained in $\varphi(A)$ because all points $p \in \phi(A)$ have the format: $\begin{pmatrix} t^2 \\ \sqrt{2}ts \\ s^2 \end{pmatrix}$, $s, t \in \mathbb{R}$ and $\sqrt{2}ts = 0 \neq 1$ for $t^2 = 0$ and $s^2 = 1$.