Machine Learning 1
Exercise 8

Group: BSSBCH

December 11, 2017

Matthias Bigalke, 339547, maku@win.tu-berlin.de
Tolga Buz, 346836, buz_tolga@yahoo.de
Alejandro Hernandez, 395678, alejandrohernandezmunuera@gmail.com
Aitor Palacios Cuesta, 396276, aitor.palacioscuesta@campus.tu-berlin.de
Christof Schubert, 344450, christof.schubert@campus.tu-berlin.de
Daniel Steinhaus, 342563, dany.steinhaus@googlemail.com

Exercise 1

Kernology

a)

i.

Show that $k(x, x') = a \qquad a \in \mathbb{R}^+$ is a Mercer kernel:

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

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j a \tag{2}$$

$$= a \sum_{i=1}^{n} c_i \sum_{j=1}^{n} c_j \tag{3}$$

$$= a (\sum_{i=1}^{n} c_i)(\sum_{j=1}^{n} c_j) \tag{4}$$

$$= a (\sum_{i=1}^{n} c_i)^2 \geq 0 \tag{5}$$

ii.

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

$$
\begin{aligned}
k(x_i, x_j) &= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j k(x_i, x_j) && (6)\\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \langle x_i, x_j \rangle && (7)\\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \ \vec{x_i}^T \cdot \vec{x_j} && (8)\\
&= (\sum_{i=1}^{n} c_i \vec{x_i})^T (\sum_{j=1}^{n} c_j \vec{x_j}) && (9)\\
&= \vec{v}^T \vec{v} = ||\vec{v}||^2 \geq 0 && (10)
\end{aligned}
$$

iii.

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

$$
\begin{aligned}
k(x_i, x_j) &= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j k(x_i, x_j) && (11)\\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j f(x_i) \cdot f(x_j) && (12)\\
&= (\sum_{i=1}^{n} c_i f(x_i))(\sum_{j=1}^{n} c_j f(x_j)) && (13)\\
&= (\sum_{i=1}^{n} c_i f(x_i))^2 \geq 0 && (14)\\
& && (15)
\end{aligned}
$$

b)

i.

Let k_1 and k_2 be two Mercer kernels. Show that:

$$
k(x, x') = k_1(x, x') + k_2(x, x') \tag{16}
$$

is also a Mercer kernel.

$$\sum_{i=1}^{d}\sum_{j=1}^{d} k(x_i, x'_j) = k_1(x, x') + k_2(x, x') \tag{17}$$

$$= \sum_{i=1}^{d}\sum_{j=1}^{d}(k_1(x_i, x'_j) + k_2(x_i, x'_j)) \tag{18}$$

$$= \sum_{i=1}^{d}\sum_{j=1}^{d} k_1(x_i, x'_j) + \sum_{i=1}^{d}\sum_{j=1}^{d} k_2(x_i, x'_j) \tag{19}$$

As both terms are non-negative (Mercer kernels), the sums is also non-negative.

ii.

Let k_1 and k_2 be two Mercer kernels. Show that:

$$k(x, x') = k_1(x, x') \cdot k_2(x, x') \tag{20}$$

is also a Mercer kernel.
The *representer theorem* says, that every kernel has a corresponding scalar product. Let $\phi : \mathbb{R}^d \mapsto \mathbb{R}^a$ be the feature maps such that $k_1(x, x') = \langle\phi(x), \phi(x')\rangle$ and for k_2 let $\psi : \mathbb{R}^d \mapsto \mathbb{R}^b$ be such that $k_2(x, x') = \langle\psi(x), \psi(x')\rangle$.

$$k(x, x') = k_1(x, x') \cdot k_2(x, x') \tag{21}$$

$$= \langle\phi(x), \phi(x')\rangle \cdot \langle\psi(x), \psi(x')\rangle \tag{22}$$

$$= \phi(x)^T \phi(x') \cdot \psi(x)^T \psi(x') \tag{23}$$

$$= \sum_{i=1}^{a} \phi_i(x)\phi_i(x') \cdot \sum_{j=1}^{b} \psi_j(x)\psi_j(x') \tag{24}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} \phi_i(x)\phi_i(x') \cdot \psi_j(x)\psi_j(x') \tag{25}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} \phi_i(x)\psi_j(x)\phi_i(x')\psi_j(x') \tag{26}$$

Now let us define another feature map $\chi(x) : \mathbb{R}^d \mapsto \mathbb{R}^{a\times b}$ to be $\chi(x) = \phi(x)^T\psi(x)$. Note that $\chi_{ij}(x) = \phi_i(x)\psi_j(x)$.

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} \chi_{ij}(x)\chi_{ij}(x') \tag{27}$$

$$= \chi(x)^T \chi(x') \tag{28}$$

The product of two Mercer kernel can be rewritten as a scalar product of the feature map χ. So the resulting product must be also a Mercer kernel.

c)

Show that $k(x,x') = (\langle x, x'\rangle + \nu)^l$ is a Mercer kernel, where $\nu \in \mathbb{R}_+$.
It can be rewritten as $k(x,x')$:

$$k_\nu(x,x') = \nu \tag{29}$$
$$k_s(x,x') = \langle x, x'\rangle \tag{30}$$
$$k_+(x,x') = k_s(x,x') + k_\nu(x,x') \tag{31}$$
$$k(x,x') = \prod^{l} k_+(x,x') \tag{32}$$

Therefore k is also a Mercer kernel.

d)

$$\begin{aligned} k(x,x') &= \exp\left(-\frac{||x-x'||^2}{2\sigma^2}\right) \\ &= \exp\left(-\frac{||x||^2 - 2\langle x', x\rangle + ||x'||^2}{2\sigma^2}\right) \\ &= \exp\left(-\frac{||x||^2}{2\sigma^2}\right) \cdot \exp\left(-\frac{||x'||^2}{2\sigma^2}\right) \cdot \exp\left(\frac{2\langle x', x\rangle}{2\sigma^2}\right) \end{aligned}$$

Remember $\exp(-\frac{||x||^2}{2\sigma^2}) \cdot \exp(-\frac{||x'||^2}{2\sigma^2}) = f(x) \cdot f(x')$ and $\frac{2\langle x', x\rangle}{2\sigma^2}$ are Mercer kernels (from a,b,c). Now we have to show that $\exp(k(x,x'))$ is a Mercer kernel, if k is a Mercer kernel.

$$\exp(k(x,x')) = \lim_{i\to\infty} \left(1 + k(x,x') + \frac{k(x,x')^2}{2} + \frac{k(x,x')^3}{6} + ... + \frac{k(x,x')^i}{i!}\right)$$

From a),b),c) we know this must also be a Mercer kernel and hence our Gaussian kernel is a Mercer kernel.

Exercise 2

The Feature Map

a)

Show that the following expression is correct:

$$\langle \phi(x), \phi(y)\rangle = \left(\sum_{i=1}^{2} x_i y_i\right)^2 \tag{33}$$

$$\left(\sum_{i=1}^{2} x_i y_i\right)^2 = (x_1y_1 + x_2y_2)^2 = \tag{34}$$

$$= x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2 \tag{35}$$

$$\tag{36}$$

Using the feature map we get:

$$\langle \phi(x), \phi(y) \rangle = \langle \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix}, \begin{pmatrix} y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \end{pmatrix} \rangle \tag{37}$$

$$= x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2 \tag{38}$$

$$\tag{39}$$

As both results are the same, ϕ and $\mathbb{R}^3$ are possible choices for feature map and feature space.

b)

We gave a mathematical description of the image set, because we are not realy sure what an explicit description of an image is.

i.

$$\left\{ \begin{pmatrix} cos^2\Theta \\ \sqrt{2}cos(\Theta)sin(\Theta) \\ sin^2\Theta \end{pmatrix} ;\ \Theta \in [0, 2\pi] \right\}$$

ii.

$$\left\{ \begin{pmatrix} t^2 \\ \sqrt{2}ts \\ s^2 \end{pmatrix} ;\ t, s \in \mathbb{R} \right\}$$

c)

We take three linear independent points from our image of b) i: $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{\sqrt{2}}{2} \\ \frac{1}{2} \end{pmatrix}$

With the solution of the previous image, we can describe the plane parametrically as $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + r \cdot \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + s \cdot \begin{pmatrix} -\frac{1}{2} \\ \frac{\sqrt{2}}{2} \\ \frac{1}{2} \end{pmatrix} ;\ r, s \in \mathbb{R} \right\}$

d)

$\begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$ is not in $\varphi(A)$, since t^2 can not take negative values for $t \in \mathbb{R}$