Machine Learning I at TU Berlin

Assignment 7 - Group PTHGL

December 9, 2018

Exercise 1: Kernology

a)

In the following sections, we show that the given kernels are Mercer kernels.

i)

$$k(x, x') = a, \quad a \in \mathbb{R}^+$$

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j a &= a \sum_{i=1}^{n} c_i \sum_{i=j}^{n} c_j \\ &= a \left(\sum_{i=1}^{n} c_i\right)^2 \geq 0\end{aligned}$$

With $c_i \in \mathbb{R}$, the square of the sum of c_j multiplied with a is always bigger than 0. Therefore $k(x, x') = a$ is a Mercer kernel.

ii)

$$k(x, x') =< x, x' >= \sum_{k=1}^{d} x_k x'_k$$

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \sum_{k=1}^{d} x_{i,k} x'_{j,k} &= \sum_{k=1}^{d}\sum_{i=1}^{n} c_i x_{i,k} \sum_{i=j}^{n} c_j x'_{j,k} \\ &= \sum_{k=1}^{d}\left(\sum_{i=1}^{n} c_i x_{i,k}\right)^2 \geq 0\end{aligned}$$

iii)

$$k(x,x') =< x, x' >= f(x) \cdot f(x'), \quad f : \mathbb{R}^d \to \mathbb{R} \text{ arbitrary continuous function}$$

$$\begin{aligned}
\sum_{i=1}^{n} \sum_{j=0}^{n} c_i c_j f(x) \cdot f(x') &= \sum_{i=1}^{n} c_i f(x_i) \sum_{i=j}^{n} c_j f(x_j) \\
&= \left(\sum_{i=1}^{n} c_i f(x_i) \right)^2 \geq 0
\end{aligned}$$

b)

k_1 and k_2 are Mercer kernels. In the following, we show that

i)

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

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

ii)

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

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k_1(x_i, x_j) \geq 0, \quad \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k_2(x_i, x_j) \geq 0 \\
&\Rightarrow \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k_1(x_i, x_j) \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k_2(x_i, x_j) \geq 0 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k_1(x_i, x_j) c_i c_j k_2(x_i, x_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i^2 c_j^2 k_1(x_i, x_j) k_2(x_i, x_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i^2 c_j^2 k(x_i, x_j) \geq 0
\end{aligned}
$$

c)

$$K(x, x') = (\langle x, x' \rangle + v)^d \quad , \quad v \in \mathbb{R}^+$$

- We know from $(a.i)$ that v is a Mercer kernel;
- And from $(a.ii)$ that $\langle x, x' \rangle$ is also a Mercer kernel;
- From $(b.i)$ we know that sum of two Mercer kernels is a Mercer kernel, so

  $$(\langle x, x' \rangle + v)$$

  is a Mercer kernel;
- From $(b.ii)$ we know that multiplication of Mercer kernels is also a Mercer kernel, so

  $$(\langle x, x' \rangle + v)^d = (\langle x, x' \rangle + v) \times (\langle x, x' \rangle + v) \times \ldots \times (\langle x, x' \rangle + v)$$

  is a Mercer kernel.

d)

$$\begin{aligned}
K(x,x') &= exp\left\{-\frac{||x-x'||^2}{2\sigma^2}\right\} \\
&= exp\left\{\frac{-||x||^2-||x'||^2+2\langle x,x'\rangle}{2\sigma^2}\right\} \\
&= exp\left\{\frac{-||x||^2-||x'||^2}{2\sigma^2}\right\} exp\left\{\frac{\langle x,x'\rangle}{\sigma^2}\right\} \\
&= C\times exp\left\{\frac{\langle x,x'\rangle}{\sigma^2}\right\}, \quad \text{where } C = exp\left\{\frac{-||x||^2-||x'||^2}{2\sigma^2}\right\} \text{is a constant} \\
&= C\times \sum_{n=0}^{\infty}\frac{\langle x,x'\rangle^n}{\sigma^2 n!}, \qquad \text{Taylor expansion of } e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}.
\end{aligned}$$

- $\dfrac{1}{\sigma^2 n!}$ is a Mercer kernel from $(a.i)$;
- $\dfrac{\langle x,x'\rangle^n}{\sigma^2 n!}$ is a Mercer kernel from $(b.ii)$;
- $\displaystyle\sum_{n=0}^{\infty}\frac{\langle x,x'\rangle^n}{\sigma^2 n!}$ is a Mercer kernel from $(b.i)$;
- $K(x,x') = exp\left\{-\dfrac{||x-x'||^2}{2\sigma^2}\right\}$ is a Mercer kernel.

Exercise 2: The Feature Map

a)

Given the polynomial Mercer Kernel of degree $d=2$, $k(x,x') = \langle x,y\rangle^2$, a valid φ needs to satisfy the following constraint: $k(x,y) = \langle\varphi(x),\varphi(y)\rangle$

With:

$$\varphi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}x_1^2\\ \sqrt{2}x_1x_2\\ x_3^2\end{pmatrix}$$

We proof:

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_3^2 \end{pmatrix}^T \begin{pmatrix} y_1^2 \\ \sqrt{2}y_1y_2 \\ y_3^2 \end{pmatrix}$$

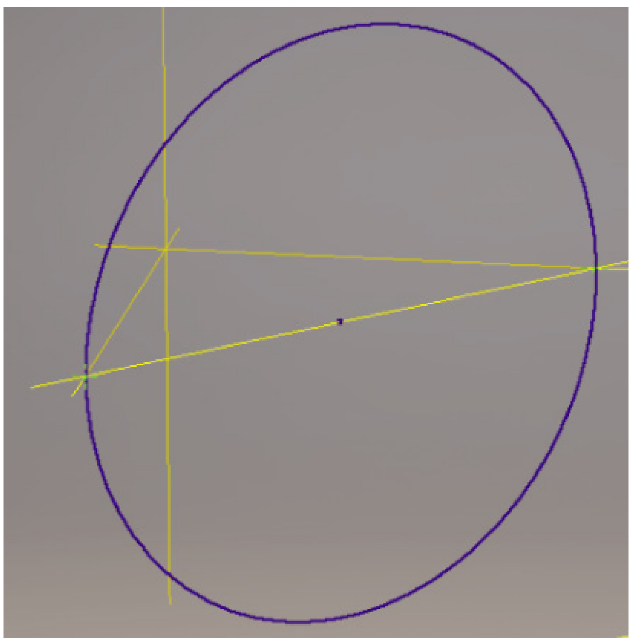
$$= x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2$$
$$= (x_1y_1 + x_2y_2)^2 = \sum_{i=1}^{2}(x_iy_i)^2 = \sum_{i=1}^{d}(x_iy_i)^2 \checkmark$$

As shown above, F and φ are suitable choices for feature space and feature map.

b.i)

The image of C under φ in F defined as follows:

$\varphi(C) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 \in [0,1], x_2 = +-\sqrt{2 \times (x_1 - x_1^2)}, x_3 = 1 - x_1\}$
Plotting the function gives us the following figure:

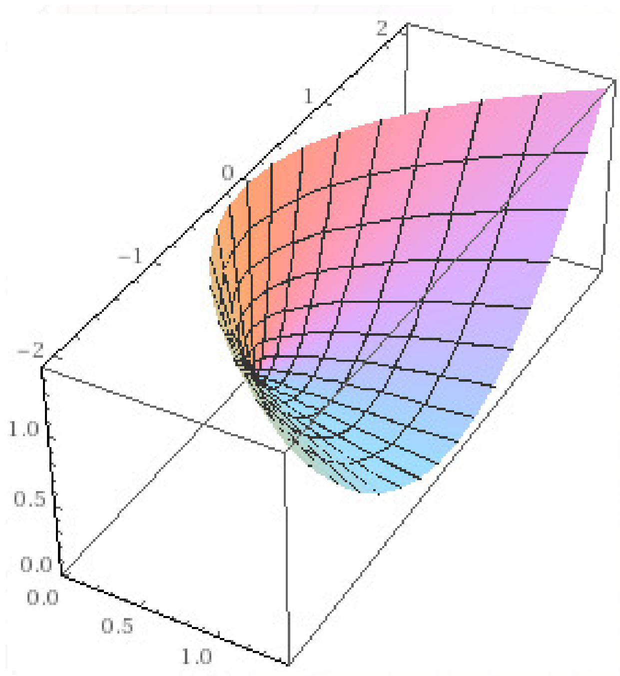


When mapped to the feature space by φ, all points are possible projected on the circle rim that is shown in the plot. The circle has a radius of $\sqrt{2}$ and lies in the plane H.

b.ii)

The image of A under φ in F defined as follows:

$\varphi(A) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1, x_3 \in \mathbb{R}^+, x_2 \in \mathbb{R}\}$ Plotting the function gives us the following figure:



When mapped to the feature space by φ, all points are possible projected on the surface that is shown in the plot. .

c)

The image of the circle $\varphi(C)$ lies on a plane H in $\mathbb{R}^3$. We can take the three points

$$\vec{p}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \vec{p}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \vec{p}_2 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

to determine the Hesse normal form of plane H, which defines th Hilbert Space.

$$H := \left(\vec{x} \cdot \begin{bmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ \frac{\sqrt{2}}{2} \end{bmatrix} \right) = \frac{\sqrt{2}}{2}$$

From the Hesse normal form, we can see that H is perpendicular to the x_1 and x_3 plane, and has a distance of $\frac{\sqrt{2}}{2}$ to the the origin.

d)

We need to find a point P in F which does not belong to the image of plane A.
We know that the image of A is described as:

$$\varphi(A) = \begin{pmatrix} t^2 \\ \sqrt{2}ts \\ s^2 \end{pmatrix}$$

The feature space F is $\mathbb{R}^3$, therefore the point P(-1,0,-1) is in F, but not in $\varphi(A)$.