> A kernel function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ must satisfy the *Mercer's condition*, which verifies that for any sequence of data points $x_1, \ldots, x_n \in \mathbb{R}^d$ 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 this 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 Hilber 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
>
> $$k(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathcal{F}} \quad \text{for all } x, x' \in \mathbb{R}^d.$$

---

1. Mercer Kernels

> (a) *Show* that the following are Mercer kernels.

> i. $k(x, x') = \langle x, x' \rangle$

We need to prove the inequality given above for an arbitrary sequence of data points $x_1, \ldots, x_n \in \mathbb{R}^d$ and coefficients $c_1, \ldots, c_n \in \mathbb{R}$, namely:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \langle x_i, x_j \rangle \geq 0$$

Using the bilinearity of the inner product we get:

$$\left\langle \sum_{i=1}^{n} c_i x_i, \sum_{j=1}^{n} c_j x_j \right\rangle \geq 0$$

Note that the two terms in the inner product are identical. Since $\langle \cdot, \cdot \rangle$ is an inner product (hence $\langle v, v \rangle \geq 0 \; \forall v \in \mathbb{R}^d$), then the inequality is satisfied.

> ii. $k(x, x') = f(x) \cdot f(x')$ where $f : \mathbb{R}^d \to \mathbb{R}$ is an arbitrary continuous function.

We need to prove, for $x_1, \ldots, x_n \in \mathbb{R}^d$ and $c_1, \ldots, c_n \in \mathbb{R}$ the following:

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

which can be rewritten as follows (since the dot product is bilinear)

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

Again, the two terms are identical. Then the above expression can be rewritten as

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

which is always true.

> (b) 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 the following are again Mercer Kernels.

> i. $k(x, x') = k_1(x, x') + k_2(x, x')$

We need to prove, for $x_1, \ldots, x_n \in \mathbb{R}^d$ and $c_1, \ldots, c_n \in \mathbb{R}$ the following:

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

which can be rewritten as (due to the distributive properties of the operations in $\mathbb{R}$)

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

Since both k_1 and k_2 are Mercer kernels, both quantities in the brackets are non-negative, therefore the inequality is true.

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

We assumed that k_1 and k_2 have a finite-dimensional feature map associated to them. Let $\varphi_1 : \mathbb{R}^d \to \mathbb{R}^N$ be the feature map of k_1 and let $\varphi_2 : \mathbb{R}^d \to \mathbb{R}^M$ be the feature map for k_2. Then we can rewrite k_1 and k_2 as follows:

$$k_1(x, x') = \langle \varphi_1(x), \varphi_1(x') \rangle = \sum_{i=1}^{N} \varphi_{1,i}(x)\varphi_{1,i}(x')$$

$$k_2(x, x') = \langle \varphi_2(x), \varphi_2(x') \rangle = \sum_{j=1}^{M} \varphi_{2,j}(x)\varphi_{2,j}(x')$$

Then we have

$$k(x, x') = k_1(x, x') \cdot k_2(x, x') = \left(\sum_{i=1}^{N} \varphi_{1,i}(x)\varphi_{1,i}(x')\right) \cdot \left(\sum_{j=1}^{M} \varphi_{2,j}(x)\varphi_{2,j}(x')\right)$$

and due to the distributive properties of operations in $\mathbb{R}$ we obtain:

$$k(x,x') = \sum_{i=1}^{N}\sum_{j=1}^{M} \varphi_{1,i}(x)\varphi_{1,i}(x')\varphi_{2,j}(x)\varphi_{2,j}(x') = \sum_{i=1}^{N}\sum_{j=1}^{M} \underbrace{(\varphi_{1,i}(x)\varphi_{2,j}(x))}_{\text{term dependent on } x} \underbrace{(\varphi_{1,i}(x')\varphi_{2,j}(x'))}_{\text{term dependent on } x'}$$

Then we can define the feature function $\Phi(x)$ as $\Phi_{i,j}(x) = \varphi_{1,i}(x)\, \varphi_{2,j}(x)$. Then we obtain

$$k(x,x') = \langle \Phi(x), \Phi(x') \rangle$$

Since $k(x,x')$ can be rewritten as an inner product in a feature space, it is a mercer kernel.

> (c) *Show* using the results above that the polynomial kernel of degree d, where $k(x,x') = (\langle x,x' \rangle + \theta)^d$ and $\theta \in \mathbb{R}^+$, is a Mercer kernel.

From exercise (a)i we know that $k_0(x,x') = \langle x,x' \rangle$ is a Mercer kernel.
From exercise (a)ii we know that $k_1(x,x') = f(x) \cdot f(x') = \sqrt{\theta} \cdot \sqrt{\theta} = \theta$ is a Mercer kernel.
From exercise (b)i we know that $k_2(x,x') = \langle x,x' \rangle + \theta$ is a Mercer kernel.
From exercise (b)ii (applied $d-1$ times) we know that $k(x,x') = (\langle x,x' \rangle + \theta)^d$ is a Mercer kernel.

---

2. The Feature Map

> 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 $\varphi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix}$ are possible choices for feature space and feature map.

We need to show that

$$\langle \varphi(x), \varphi(y) \rangle_{\mathbb{R}^3} = k(x,y) = \langle x,y \rangle^2_{\mathbb{R}^2} \quad \forall x,y \in \mathbb{R}^2$$

We have

$$\langle \varphi(x), \varphi(y) \rangle_{\mathbb{R}^3} = x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2$$
$$\langle x,y \rangle^2_{\mathbb{R}^2} = (x_1y_1 + x_2y_2)^2 = x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2$$

Thus the equality is respected and the choice is possible.

> (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 $\mathbb{R}^3$.

For $0 \le \theta < 2\pi$ we have:

$$\begin{aligned} \varphi(C) &= \varphi \left\{ \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} : 0 \le \theta < 2\pi \right\} \\ &= \left\{ \begin{pmatrix} (\cos\theta)^2 \\ \sqrt{2}(\cos\theta)(\sin\theta) \\ (\sin\theta)^2 \end{pmatrix} : 0 \le \theta < 2\pi \right\} \end{aligned}$$

Note that

$$\varphi(C) \subseteq H := \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : x + z = 1 \right\}$$

H is a plane in $\mathbb{R}^3$ because fixes two coordinates (x, z) on the same straight line and leaves the third coordinate (y) uncostrained on $\mathbb{R}$.

> (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 $\varphi(A)$.

Since $\varphi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^2 \\ \sqrt{2} x_1 x_2 \\ x_2^2 \end{pmatrix}$, the first coordinate will be always non-negative for all $x_1 \in \mathbb{R}$, thus we can simply take a point where the first coordinate is negative, for example $P = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$

> (d) *Find* a feature map associated to the kernel $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ with $k(x, y) = \langle x, y \rangle^2 = \left(\sum_{i=1}^d x_i y_i \right)^2$.

We have

$$\begin{aligned} k(x, y) = \left(\sum_{i=1}^d x_i y_i \right)^2 &= \left(\sum_{i=1}^d x_i y_i \right) \left(\sum_{j=1}^d x_j y_j \right) = \sum_{i=1}^d \sum_{j=1}^d x_i y_i x_j y_j \\ &= \sum_{i=1}^d (x_i y_i)^2 + \sum_{i \ne j} x_i x_j y_i y_j \\ &= \sum_{i=1}^d (x_i y_i)^2 + \sum_{1 \le i < j \le d} 2 x_i x_j y_i y_j \end{aligned}$$

We want to have the same result as in (a), so we construct $\varphi(x)$ such that $\langle \varphi(x), \varphi(x) \rangle^2$

equals the expression above. To do that, we can choose

$$\varphi\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} = \begin{pmatrix} x_1^2 \\ \vdots \\ x_d^2 \\ \sqrt{2}x_1x_2 \\ \sqrt{2}x_1x_3 \\ \vdots \\ \sqrt{2}x_1x_d \\ \sqrt{2}x_2x_3 \\ \vdots \\ \sqrt{2}x_{d-1}x_d \end{pmatrix}$$