

Machine Learning

1. Kernology

(a)

$$\begin{aligned}
 \text{i. } & \sum_{i=1}^n \sum_{j=1}^n C_i C_j k(x_i, x_j) \\
 &= a(c_1(c_1 + c_2 + \dots + c_n) + c_2(c_1 + c_2 + \dots + c_n) + \dots + c_n(c_1 + c_2 + \dots + c_n)) \\
 &= a(c_1 + c_2 + \dots + c_n)(c_1 + c_2 + \dots + c_n) \\
 &= a(c_1 + c_2 + \dots + c_n)^2 \geq 0
 \end{aligned}$$

Therefore $k(x, x') = a$ is a Mercer kernel.

ii. According to the definition of inner product, we can get the following properties:

· Conjugate symmetry:

$$\langle x, y \rangle = \overline{\langle y, x \rangle};$$

· Linearity in the first argument:

$$\langle ax, y \rangle = a \langle x, y \rangle;$$

$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle;$$

· Positive-definiteness:

$$\langle x, x \rangle \geq 0.$$

When $F = \mathbb{R}$, the following equations are also satisfied:

$$\langle x, ay \rangle = \overline{\langle ay, x \rangle} = \bar{a} \overline{\langle y, x \rangle} = \bar{a} \langle x, y \rangle;$$

$$\langle x, y + z \rangle = \overline{\langle y + z, x \rangle} = \overline{\langle y, x \rangle + \langle z, x \rangle} = \langle x, y \rangle + \langle x, z \rangle.$$

Therefore:

$$\begin{aligned}
 & \sum_{i=1}^n \sum_{j=1}^n c_i c_j k(x_i, x_j) \\
 &= c_1^2 \langle x_1, x_1 \rangle + c_1 c_2 \langle x_1, x_2 \rangle + \dots + c_1 c_n \langle x_1, x_n \rangle + c_2 c_1 \langle x_2, x_1 \rangle + c_2^2 \langle x_2, x_2 \rangle \\
 &+ \dots + c_2 c_n \langle x_2, x_n \rangle + \dots + c_n c_1 \langle x_n, x_1 \rangle + c_n c_2 \langle x_n, x_2 \rangle + \dots + c_n^2 \langle x_n, x_n \rangle \\
 &= c_1^2 \langle x_1, x_1 \rangle + \frac{x_2}{x_1} c_1 c_2 \langle x_1, x_2 \rangle + \dots + \frac{x_n}{x_1} c_1 c_n \langle x_1, x_n \rangle + \frac{x_2}{x_1} c_2 c_1 \langle x_2, x_1 \rangle + \left(\frac{x_2}{x_1}\right)^2 c_2^2 \langle x_2, x_2 \rangle \\
 &+ \dots + \frac{x_2}{x_1} \cdot \frac{x_n}{x_1} c_2 c_n \langle x_2, x_n \rangle + \dots + \frac{x_n}{x_1} c_n c_1 \langle x_n, x_1 \rangle + \frac{x_n}{x_1} \cdot \frac{x_2}{x_1} c_n c_2 \langle x_n, x_2 \rangle + \dots + \left(\frac{x_n}{x_1}\right)^2 c_n^2 \langle x_n, x_n \rangle \\
 &= \langle x_1, x_1 \rangle \cdot \left(c_1 \left(c_1 + \frac{x_2}{x_1} c_2 + \frac{x_3}{x_1} c_3 + \dots + \frac{x_n}{x_1} c_n \right) + \frac{x_2}{x_1} c_2 \left(c_1 + \frac{x_2}{x_1} c_2 + \frac{x_3}{x_1} c_3 + \dots + \frac{x_n}{x_1} c_n \right) + \dots + \frac{x_n}{x_1} c_n \left(c_1 + \frac{x_2}{x_1} c_2 + \frac{x_3}{x_1} c_3 + \dots + \frac{x_n}{x_1} c_n \right) \right)
 \end{aligned}$$

$$\begin{aligned}
&= \langle x_1, x_1 \rangle \cdot (c_1 + \frac{x_2}{x_1}c_2 + \dots + \frac{x_n}{x_1}c_n) \cdot (c_1 + \frac{x_2}{x_1}c_2 + \dots + \frac{x_n}{x_1}c_n) \\
&= \langle x_1, x_1 \rangle \cdot (c_1 + \frac{x_2}{x_1}c_2 + \dots + \frac{x_n}{x_1}c_n)^2 \geq 0.
\end{aligned}$$

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

$$\begin{aligned}
\text{iii. } &\sum_{i=1}^n \sum_{j=1}^n c_i c_j k(x_i, x_j) \\
&= c_1^2 f^2(x_1) + c_1 c_2 f(x_1) f(x_2) + \dots + c_1 c_n f(x_1) f(x_n) + c_2 c_1 f(x_2) f(x_1) + c_2^2 f^2(x_2) + \dots + c_2 c_n f(x_2) f(x_n) + \\
&\dots + c_n c_1 f(x_n) f(x_1) + c_n c_2 f(x_n) f(x_2) + \dots + c_n^2 f^2(x_n) \\
&= c_1 f(x_1) (c_1 f(x_1) + c_2 f(x_2) + \dots + c_n f(x_n)) + c_2 f(x_2) (c_1 f(x_1) + c_2 f(x_2) + \dots + c_n f(x_n)) + \dots + \\
&c_n f(x_n) (c_1 f(x_1) + c_2 f(x_2) + \dots + c_n f(x_n)) \\
&= (c_1 f(x_1) + c_2 f(x_2) + \dots + c_n f(x_n)) \cdot (c_1 f(x_1) + c_2 f(x_2) + \dots + c_n f(x_n)) \\
&= (c_1 f(x_1) + c_2 f(x_2) + \dots + c_n f(x_n))^2 \geq 0.
\end{aligned}$$

Therefore $k(x, x') = f(x) \cdot f(x')$ is a Mercer kernel.

(b)

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

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

$$\text{ii. as } \sum_{i=1}^n \sum_{j=1}^n a_i a_j k_1(x_i, x_j) \geq 0, \sum_{f=1}^n \sum_{g=1}^n b_f b_g k_2(x_f, x_g) \geq 0$$

holds for any coefficients a_i, a_j, b_f, b_g ,

$$\begin{aligned}
&\text{so, } (\sum_{i=1}^n \sum_{j=1}^n a_i a_j k_1(x_i, x_j)) \cdot (\sum_{f=1}^n \sum_{g=1}^n b_f b_g k_2(x_f, x_g)) \\
&= \sum_{i=1}^n \sum_{j=1}^n \sum_{f=1}^n \sum_{g=1}^n a_i a_j b_f b_g (k_1(x_i, x_j) k_2(x_f, x_g)) \geq 0,
\end{aligned}$$

also holds for any coefficients a_i, a_j, b_f, b_g ,

we set :

$$\begin{cases} b_f = 1, b_g = 1 & f = i, g = j \\ b_f = 0, b_g = 0 & f \neq i, \text{ or } g \neq j \end{cases}$$

so,

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{f=1}^n \sum_{g=1}^n a_i a_j b_f b_g (k_1(x_i, x_j) k_2(x_f, x_g))$$

$$= \sum_{i=1}^n \sum_{j=1}^n a_i a_j k_1(x_i, x_j) k_2(x_i, x_j) \geq 0.$$

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

- (c) $k_1(x, x') = \langle x, x' \rangle \dots \dots \dots$ a Mercer kernel as (a)ii has proved.
 $k_2(x, x') = \theta \dots \dots \dots$ a Mercer kernel as (a)i has proved.
 $k_3(x, x') = k_1(x, x') + k_2(x, x') = \langle x, x' \rangle + \theta \dots \dots \dots$ a Mercer kernel as (b)i has proved.
 $k_4(x, x') = k_3(x, x') k_3(x, x') = (\langle x, x' \rangle + \theta)^2 \dots \dots \dots$ a Mercer kernel as (b)ii has proved.
 $k_5(x, x') = k_4(x, x') k_3(x, x') = (\langle x, x' \rangle + \theta)^3 \dots \dots \dots$ a Mercer kernel as (b)ii has proved.
 \vdots
 $k_{d+2}(x, x') = k_{d+1}(x, x') k_3(x, x') = (\langle x, x' \rangle + \theta)^d \dots \dots \dots$ a Mercer kernel as (b)ii has proved.
Therefore $k(x, x') = (\langle x, x' \rangle + \theta)^d$ is a Mercer kernel.

- (d) It can be taken as known that: $\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$,

Hence $\exp(k(x, x'))$ is an infinite sum of kernels. Using (a)i and (b)i, we can get that:

$\exp(k(x, x'))$ should be Mercer kernels, if $k(x, x')$ is Mercer kernel.

as $k(x, x') = -\frac{xx'}{2\sigma}$ is a Mercer kernel, so $\exp(-\frac{xx'}{2\sigma})$ should also be a Mercer kernel.

by (a)iii: let $f(x) = \exp(-\frac{x^2}{2\sigma})$

$k(x, x') = f(x) \cdot f(x') = \exp(-\frac{x^2}{2\sigma}) \cdot \exp(-\frac{x'^2}{2\sigma})$ should be Mercer kernel.

so, $k(x, x') = \exp(-\frac{(x-x')^2}{2\sigma}) = \exp(-\frac{x^2}{2\sigma}) \cdot \exp(-\frac{x'^2}{2\sigma}) \cdot \exp(-\frac{xx'}{2\sigma})$ is also a Mercer kernel.

2. The Feature Map

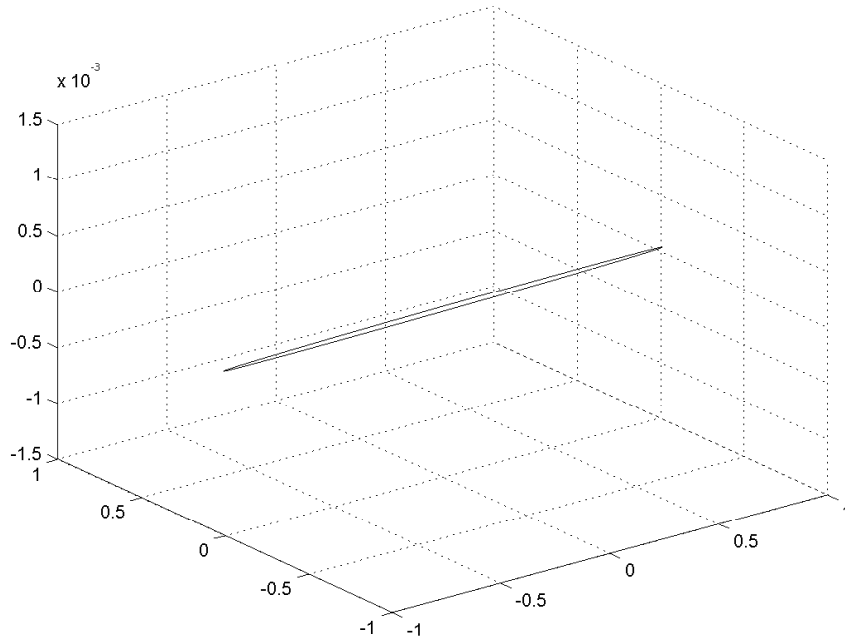
- (a) $k(x, y) = (\sum_{i=1}^2 x_i y_i) = (x_i + y_i)^2 = x_1^2 y_1^2 + x_2^2 + y_2^2 + 2x_1 y_1 x_2 y_2$
- $$\varphi_{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} \cdot \varphi_{\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}} = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1 x_2 \\ x_2^2 \end{pmatrix} \cdot \begin{pmatrix} y_1^2 \\ \sqrt{2}y_1 y_2 \\ y_2^2 \end{pmatrix}$$
- $$= x_1^2 y_1^2 + 2x_1 x_2 y_1 y_2 + x_2^2 y_2^2$$
- $$= k(x, y).$$

Therefore, F and φ are possible choices for feature space and feature map.

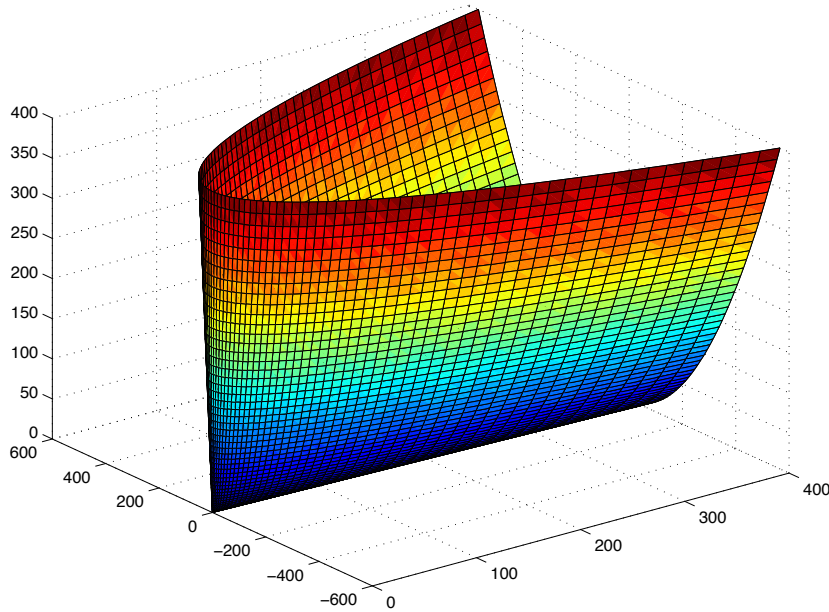
- (b)

- i. Under φ the circle maps to a closed curve which looks like a narrow ellipse whose major axe is much

longer than the minor axe. The following is the graph of function calculated by matlab:



ii. Under φ the plane maps to a curved surface which looks like an open briefcase. The following is the graph of function calculated by matlab:



(c) The plane passes three points $(1, 0, 1)$, $(0, 0, 1)$, $(\frac{1}{2}, \frac{\sqrt{2}}{2}, \frac{1}{2})$. Hence the new plane is vertical to plane formed by $X - Z$ axes.

(d) Point $(0, 1, 0)$ is not contained in $\varphi(A)$ because we cannot find the corresponding x_1, x_2 that satisfies

$$\begin{cases} x_1^2 = 0 \\ \sqrt{2}x_1x_2 = 1 \\ x_2^2 = 0 \end{cases}$$