Support Vector Machines - IV

Prof. Dan A. Simovici

UMB
1. Eigenvalues of Positive Definite Matrices

2. Hilbert Spaces

3. Kernels

4. Functions of Positive Type

5. Examples of Positive Definite Kernels
Recall that a matrix $A \in \mathbb{R}^{n \times n}$ is positive definite if $x'Ax > 0$ for $x \neq 0$.

**Theorem**

The eigenvalues of a real symmetric positive matrix are positive.

**Proof:** The eigenvalues of real symmetric matrices are real. If $\lambda$ is an eigenvalue of $A$ with the eigenvector $x$, then $Ax = \lambda x$, hence $x'Ax = \lambda x'x = \lambda \| x \|^2 > 0$. Thus, $\lambda > 0$. 
Theorem

If the eigenvalues of a real symmetric matrix are positive, then \( A \) is positive definite.

Proof: For a real symmetric matrix there exists an orthogonal matrix \( Q \) such that \( Q'AQ = D \), where

\[
D = \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
& & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_n
\end{pmatrix}
\]

If \( x \neq 0_n \), then \( x'Ax = x'Q'DQx = y'Dy \), where \( y = Qx \).

Then, \( y'Dy = \lambda_1y_1^2 + \cdots + \lambda_ny_n^2 > 0 \) because \( y = Q'x \) is a non-zero vector. Here we used the fact that \( Q^{-1} = Q' \).
Hilbert space, named after David Hilbert, generalize the notion of Euclidean space. They extend the methods of vector algebra and calculus from the two-dimensional Euclidean plane and three-dimensional space to spaces with any finite or infinite number of dimensions.
An inner product \((x, y)\) defined on a linear space \(H\) generates a norm 
\[\| x \| = \sqrt{(x, x)}.\]

A norm on a linear space generates a distance (a metric) 
\[d(x, y) = \| x - y \|.\] Thus, every normed space becomes a metric space.

A Cauchy sequence in a metric space is a sequence \((x_n)\) such that for every \(\epsilon > 0\) there exists a number \(n_\epsilon\) such that \(m, p > n_\epsilon\) imply 
\[d(x_m, x_p) < \epsilon.\]

A metric space is complete if every Cauchy sequence has a limit in that space.
What is a Hilbert Space?

Hilbert spaces are generalizations of Euclidean spaces. A Hilbert space is a linear space that is equipped with an inner product such that the metric space generated by the inner product is complete. As above, the inner product of two elements $x, y$ of a Hilbert space $H$ is denoted by $(x, y)$. Note that in the case of $\mathbb{R}^n$ (which is a special case of a Hilbert space) the inner product of $x, y$ was denoted by $x'y$. 
Example

The Euclidean space $\mathbb{R}^n$ equipped with the inner product

$$(x, y) = x_1y_1 + \cdots + x_ny_n$$

is a Hilbert space.
Example

The space $\ell^2$ that consists of infinite sequences of the form $z = (z_1, z_2, \ldots)$ such that the series $\sum_n |z_n|^2$ converges is a Hilbert space, where the inner product is defined as

$$\langle z, w \rangle = \sum_{n=1}^{\infty} z_n \overline{w_n}.$$
Example

For two functions $f, g$ such that $\int_a^b f^2(x) \, dx$ and $\int_a^b g^2(x) \, dx$ exist, an inner product can be defined as

$$(f, g) = \int_a^b f(x)g(x) \, dx.$$ 

The resulting linear space is a Hilbert space.
Definition

A kernel over $\mathcal{X}$ is a function $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ such that there exists a function $\Phi : \mathcal{X} \rightarrow H$ that satisfies the condition

$$K(u, v) = \langle \Phi(u), \Phi(v) \rangle,$$

where $H$ is a Hilbert space called the feature space.
Recall the general form of the dual optimization problem for SVMs:

$$\text{maximize for } a \sum_{i=1}^{m} a_i - \frac{1}{2} a_i a_j y_i y_j x_i' x_j$$
subject to $0 \leq a_i \leq C$ and $\sum_{i=1}^{m} a_i y_i = 0$
for $1 \leq i \leq m$.

Note the presence of the inner product $x_i' x_j$. This is replaced by the inner product $(\Phi(x_i), \Phi(x_j))$, in the Hilbert feature space, that is, by $K(x_i, x_j)$, where $K$ is a suitable kernel function.
A More General SVM Formulation

\[
\text{maximize for } a \quad \sum_{i=1}^{m} a_i - \frac{1}{2} a_i a_j y_i y_j K(x_i, x_j)
\]

\[
\text{subject to } 0 \leq a_i \leq C \text{ and } \sum_{i=1}^{m} a_i y_i = 0
\]

\[
\text{for } 1 \leq i \leq m.
\]

The hypothesis returned by the SVM algorithm is now

\[
h(x) = \text{sign} \left( \sum_{i=1}^{m} a_i y_i K(x_i, x) + b \right).
\]

with \( b = y_i - \sum_{j=1}^{m} a_j y_j K(x_j, x_i) \) for any \( x_i \) with \( 0 < a_i < C \).

Note that we do not work with the feature mapping \( \Phi \); instead we use the kernel only!
Definition

Let $S$ be a non-empty set. A function $K : S \times S \rightarrow \mathbb{C}$ is of positive type if for every $n \geq 1$ we have:

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

for every $a_i \in \mathbb{C}$ and $x_i \in S$, where $1 \leq i \leq n$.

$K : S \times S \rightarrow \mathbb{R}$ is of positive type if for every $n \geq 1$ we have

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

for every $a_i \in \mathbb{R}$ and $x_i \in S$, where $1 \leq i \leq n$. 
If $K : S \times S \rightarrow \mathbb{C}$ is of positive type, then taking $n = 1$ we have

$$aK(x, x)\bar{a} = K(x, x)|a|^2 \geq 0$$

for every $a \in \mathbb{C}$ and $x \in S$. This implies $K(x, x) \geq 0$ for $x \in S$.

Note that $K : S \times S \rightarrow \mathbb{C}$ is of positive type if for every $n \geq 1$ and for every $x_1, \ldots, x_s$ the matrix $A_{n,K}(x_1, \ldots, x_n) = (K(x_i, x_j))$ is positive definite, and, therefore it has positive eigenvalues.
Example

The function $K : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ given by $K(x, y) = \cos(x - y)$ is of positive type because

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i K(x_i, x_j) \overline{a_j} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \cos(x_i - x_j) \overline{a_j}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i (\cos x_i \cos x_j + \sin x_i \sin x_j) \overline{a_j}$$

$$= \left| \sum_{i=1}^{n} a_i \cos x_i \right|^2 + \left| \sum_{i=1}^{n} a_i \sin x_i \right|^2.$$

for every $a_i \in \mathbb{C}$ and $x_i \in S$, where $1 \leq i \leq n$. 
Definition

Let $S$ be a non-empty set. A complex-valued function $K : S \times S \rightarrow \mathbb{C}$ is *Hermitian* if $K(x, y) = \overline{K(y, x)}$ for every $x, y \in S$. 
Theorem

Let $H$ be a Hilbert space, $S$ be a non-empty set and let $f : S \to H$ be a function. The function $K : S \times S \to \mathbb{C}$ defined by

$$K(s, t) = (f(s), f(t))$$

is of positive type.
Proof

We can write

\[ \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a_j} K(t_i, t_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a_j} (f(t_i), f(t_j)) = \left\| \sum_{i=1}^{n} a_i f(a_i) \right\|^2 \geq 0, \]

which means that \( K \) is of positive type.
Theorem

Let $S$ be a set and let $F : S \times S \rightarrow \mathbb{C}$ be a positive type function. The following statements hold:

1. $F(x, y) = F(y, x)$ for every $x, y \in S$, that is, $F$ is Hermitian;
2. $F$ is a positive type function;
3. $|F(x, y)|^2 \leq F(x, x)F(y, y)$. 

Proof

Take $n = 2$ in the definition of positive type functions. We have

$$a_1 \overline{a_1} F(x_1, x_1) + a_1 \overline{a_2} F(x_1, x_2) + a_2 \overline{a_1} F(x_2, x_1) + a_2 \overline{a_2} F(x_2, x_2) \geq 0, \quad (1)$$

which amounts to

$$|a_1|^2 F(x_1, x_1) + a_1 \overline{a_2} F(x_1, x_2) + a_2 \overline{a_1} F(x_2, x_1) + |a_2|^2 F(x_2, x_2) \geq 0,$$

By taking $a_1 = a_2 = 1$ we obtain

$$p = F(x_1, x_1) + F(x_1, x_2) + F(x_2, x_1) + F(x_2, x_2) \geq 0,$$

where $p$ is a positive real number.

Similarly, by taking $a_1 = i$ and $a_2 = 1$ we have

$$q = -F(x_1, x_1) + iF(x_1, x_2) - iF(x_2, x_1) + F(x_2, x_2) \geq 0,$$

where $q$ is a positive real number.
Thus, we have

\[ F(x_1, x_2) + F(x_2, x_1) = p - F(x_1, x_1) - F(x_2, x_2), \]
\[ iF(x_1, x_2) - iF(x_2, x_1) = q + F(x_1, x_1) - F(x_2, x_2). \]

These equalities imply

\[ 2F(x_1, x_2) = P - iQ, \]
\[ 2F(x_2, x_1) = P + iQ, \]

where \( P = p - F(x_1, x_1) - F(x_2, x_2) \) and \( Q = q + F(x_1, x_1) - F(x_2, x_2) \), which shows the first statement holds.
The second part of the theorem follows by applying the conjugation in the equality of Definition.

For the final part, note that if \( F(x_1, x_2) = 0 \) the desired inequality holds immediately. Therefore, assume that \( F(x_1, x_2) \neq 0 \) and take \( a_1 = a \in \mathbb{R} \) and to \( a_2 = F(x_1, x_2) \). We have

\[
a^2 F(x_1, x_1) + a F(x_1, x_2) F(x_1, x_2) + F(x_1, x_2) a F(x_2, x_1) + F(x_1, x_2) F(x_1, x_2) F(x_2, x_2) \geq 0,
\]

which amounts to

\[
a^2 F(x_1, x_1) + 2a |F(x_1, x_2)| + |F(x_1, x_2)|^2 F(x_2, x_2) \geq 0.
\]

If \( F(x_1, x_1) \) this trinomial in \( a \) must be non-negative for every \( a \), which implies

\[
|F(x_1, x_2)|^4 - |F(x_1, x_2)|^2 F(x_1, x_1) F(x_2, x_2) \leq 0.
\]

Since \( F(x_1, x_2) \neq 0 \), the desired inequality follows.
Theorem

A real-valued function $G : S \times S \rightarrow \mathbb{R}$ is a positive type function if it is symmetric and
\[
\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j G(x_i, x_j) \geq 0
\] (2)

for $a_1, \ldots, a_n \in \mathbb{R}$ and $x_1, \ldots, x_n \in S$.

In other words $G$ is a positive type function iff $(G(x_i, x_j))$ is a positive-definite matrix for any $x_1, \ldots, x_n \in S$. 

Theorem

Let $S$ be a non-empty set. If $K_i : S \times S \rightarrow \mathbb{C}$ for $i = 1, 2$ are functions of positive type, then their pointwise product $K_1 K_2$ defined by $(K_1 K_2)(x, y) = K_1(x, y) K_2(x, y)$ is of positive type.
Proof

Since $K_i$ is a function of positive type, the matrix

$$A_{n,K_i}(x_1,\ldots,x_n) = (K_i(x_j, x_h))$$

is positive, where $i = 1, 2$. Thus, such matrices can be factored as

$$A_{n,K_1}(x_1,\ldots,x_n) = P^H P \text{ and } A_{n,K_2}(x_1,\ldots,x_n) = R^H R$$

for $i = 1, 2$. Therefore, we have:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i K_1(x_i, x_j) K_2(x_i, x_j) \overline{a_j}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i K(x_i, x_j) \cdot \left( \sum_{m=1}^{n} \overline{r_{mi}} r_{mj} \right) \overline{a_j}$$

$$= \sum_{m=1}^{n} \left( \sum_{i=1}^{n} a_i \overline{r_{mi}} \right) K(x_i, x_j) \left( \sum_{j=1}^{n} r_{jm} \overline{a_j} \right) \geq 0,$$

which shows that $(K_1 K_2)(x, y)$ is a function of positive type.
Theorem

Let $S$ be a non-empty set. The set of functions of positive type is closed with respect to multiplication with non-negative scalars and with respect to addition.
Which of the following functions are kernels?

For \( x, y \in \mathbb{R}^n \):

\[
K(x, y) = \sum_{i=1}^{n} (x_i + y_i)
\]

\( K \) is not a kernel. Indeed, for \( x = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \) and \( y = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \) we have

\( k_{11} = K(x, x) = 2, \ k_{12} = K(x, y) = 3 = k_{21}, \) and \( k_{22} = K(y, y) = 4. \)

The matrix of \( K \) is

\[
\begin{pmatrix}
k_{11} & k_{12} \\
k_{21} & k_{22}
\end{pmatrix}
= \begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}.
\]
Its characteristic polynomial is

\[
\det \begin{pmatrix} 2 - \lambda & 3 \\ 3 & 4 - \lambda \end{pmatrix} = \lambda^2 - 6\lambda - 1.
\]

and has a negative eigenvalue.
\[ K_2(x, y) = \prod_{j=1}^{n} h \left( \frac{x_i - c}{a} \right) h \left( \frac{y_i - c}{a} \right), \]

where \( h(x) = \cos(1.75x) e^{-\frac{x^2}{2}} \).

\( K_2 \) is a kernel because it can be written as a product \( K_2 = f(x)f(y) \).
Functions of Positive Type

\[ K_3(x, y) = -\frac{(x, y)}{\|x\| \|y\|} \]

\( K_3 \) is not a kernel because it has negative eigenvalues.
$K_4(x, y) = \sqrt{\|x - y\|^2 + 1}$

$K_4$ is not a kernel. Indeed, for $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ the matrix

$$
\begin{pmatrix}
  k_{11} & k_{12} \\
  k_{21} & k_{22}
\end{pmatrix} = \begin{pmatrix} 1 & 5 \\ 5 & 1 \end{pmatrix}
$$

has a negative eigenvalue.
Example

A special case of functions of positive type on $\mathbb{R}^n$ are obtained by defining $K : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ as $K_f(x, y) = f(x - y)$, where $f : \mathbb{R}^n \rightarrow \mathbb{C}$ is a continuous function on $\mathbb{R}^n$. $K$ is translation invariant and is designated as a **stationary kernel**.
A function $K : S \times S \rightarrow \mathbb{C}$ defined by $K(s, t) = (f(s), f(t))$, where $f : S \rightarrow H$ is of positive type, where $H$ is a Hilbert space.

The reverse is also true:
If $K$ is of positive type a special Hilbert space exists such that $K$ can be expressed as an inner product on this space (Aronszajn’s Theorem).

This fact is essential for data kernelization that is essential for support vector machines.
Theorem

(Aronszajn’s Theorem) Let $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be a positive type kernel. Then, there exists a Hilbert space $H$ of functions and a feature mapping $\Phi : \mathcal{X} \rightarrow H$ such that $K(x, y) = (\Phi(x), \Phi(y))$ for all $x, y \in \mathcal{X}$. Furthermore, $H$ has the reproducing property which means that for every $h \in H$ we have

$$h(x) = (h, K(x, \cdot)).$$

The function space $H$ is called a reproducing Hilbert space associated with $K$. 
Definition

A continuous linear operator on a Hilbert space $H$ is **positive** if $(h(x), x)) \geq 0$ for every $x \in H$.

$h$ is **positive definite** if it is positive and invertible.

If $h$ is an operator on a space of functions and $h(f)$ is the function defined as $h(f)(x) = \int K(x, y)f(y) \, dy$, then we say that $K$ is the kernel of $h$. 
Theorem

(Mercer’s Theorem) Let \( K : [0, 1] \times [0, 1] \to \mathbb{R} \) be a function continuous in both variables that is the kernel of a positive operator \( h \) on \( L^2([0, 1]) \). If the eigenfunctions of \( h \) are \( \phi_1, \phi_2, \ldots \) and they correspond to the eigenvalues \( \mu_1, \mu_2, \ldots \), respectively then we have:

\[
K(x, y) = \sum_{j=1}^{\infty} \mu_j \phi_j(x) \overline{\phi_j(y)},
\]

where the series \( \sum_{j=1}^{\infty} \mu_j \phi_j(x) \overline{\phi_j(y)} \) converges uniformly and absolutely to \( K(x, y) \).
From the equality for the kernel of a positive operator

\[ K(u, v) = \sum_{n=0}^{\infty} a_n \phi_n(u) \phi_n(v) \]

with \( a_n > 0 \) we can construct a mapping \( \Phi \) into a feature space (in this case the potentially infinite \( \ell_2 \)) as

\[ \Phi(u) = \sum_{n=0}^{\infty} \sqrt{a_n} \phi_n(u). \]
Example

For $c > 0$ a polynomial kernel of degree $d$ is the kernel defined over $\mathbb{R}^n$ by

$$K(u, v) = (u'v + c)^d.$$  

As an example, consider $n = 2$, $d = 2$ and the kernel $K(u, v) = (u'v + c)^2$. We have:

$$K(u, v) = (u_1v_1 + u_2v_2 + c)^2$$
$$= u_1^2v_1^2 + u_2^2v_2^2 + c^2 + 2u_1v_1u_2v_2 + 2u_1v_1c + 2u_2v_2c,$$
Example (cont’d)

Feature space is $\mathbb{R}^6$

$$K(u, v) = \begin{pmatrix} u_1^2 & \sqrt{2u_1u_2} & \sqrt{2c} & \sqrt{2}c \\ u_2^2 & \sqrt{2c} & \sqrt{2} & c \end{pmatrix}' \begin{pmatrix} v_1^2 \\ \sqrt{2v_1v_2} \\ \sqrt{2cv_1} \\ \sqrt{2}cv_2 \\ c \end{pmatrix} = \Phi(u)'\Phi(v) \text{ and } \Phi(x) = \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2x_1x_2} \\ \sqrt{2cx_1} \\ \sqrt{2cx_2} \\ c \end{pmatrix}$$
In general, features associated to a polynomial kernel of degree \( d \) are all monomials of degree \( d \) associated to the original features. It is possible to show that polynomial kernels of degree \( d \) on \( \mathbb{R}^n \) map the input space to a space of dimension \( \binom{n+d}{d} \).
For the kernel $K(u, v) = (u'v + 1)^2$ we have

$$
\Phi \left( \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix}
\frac{x_1^2}{\sqrt{2x_1}} \\
\frac{x_2^2}{\sqrt{2x_2}} \\
\sqrt{2x_1} \\
\sqrt{2x_2} \\
1
\end{pmatrix}.
$$
Examples of Positive Definite Kernels

For the kernel \( K(u, v) = (u'v + 1)^2 \) we have

\[
\begin{align*}
\Phi \left( \begin{array}{c} 1 \\ 1 \\ \sqrt{2} \\ \sqrt{2} \\ 1 \\ \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \end{array} \right), & \Phi \left( \begin{array}{c} 1 \\ 1 \\ \sqrt{2} \\ \sqrt{2} \\ 1 \\ \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \end{array} \right), \\
\Phi \left( \begin{array}{c} 1 \\ 1 \\ \sqrt{2} \\ \sqrt{2} \\ 1 \\ \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \end{array} \right), & \Phi \left( \begin{array}{c} 1 \\ 1 \\ \sqrt{2} \\ \sqrt{2} \\ 1 \\ \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \end{array} \right), \\
\Phi \left( \begin{array}{c} 1 \\ 1 \\ \sqrt{2} \\ \sqrt{2} \\ 1 \\ \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \end{array} \right)
\end{align*}
\]

For this set of points differences occur in the third, fourth, and fifth features.
Definition

To any kernel $K$ we can associate a normalized kernel $K'$ defined by

$$K'(u, v) = \begin{cases} 0 & \text{if } K(u, u) = 0 \text{ or } K(v, v) = 0, \\ \frac{K(u, v)}{\sqrt{K(u, u)} \sqrt{K(v, v)}} & \text{otherwise}. \end{cases}$$

If $K(u, u) \neq 0$, then $K'(u, u) = 1$. 
Theorem

Let $K$ be a positive type kernel. For any $u, v \in \mathcal{X}$ we have

$$K(u, v)^2 \leq K(u, u)K(v, v).$$

**Proof:** Consider the matrix

$$K = \begin{pmatrix} K(u, u) & K(u, v) \\ K(v, u) & K(v, v) \end{pmatrix}$$

$K$ is positive, so its eigenvalues $\lambda_1, \lambda_2$ must be non-negative. Its characteristic equation is

$$\begin{vmatrix} K(u, u) - \lambda & K(u, v) \\ K(v, u) & K(v, v) - \lambda \end{vmatrix} = 0$$

Equivalently,

$$\lambda^2 - (K(u, u) + K(v, v))\lambda + \det(K) = 0$$

Therefore, $$\lambda_1 \lambda_2 = \det(K) \geq 0$$ and this implies

$$K(u, u)K(v, v) - K(u, v)^2 \geq 0.$$
Theorem

Let $K$ be a positive type kernel. Its normalized kernel is a positive type kernel.

Proof: Let $\{x_1, \ldots, x_m\} \subseteq \mathcal{X}$ and $c \in \mathbb{R}^m$. We prove that

$$\sum_{i,j} c_i c_j K'(x_i, x_j) \geq 0.$$

If $K(x_i, x_i) = 0$, then $K(x_i, x_j) = 0$ and, thus, $K'(x_i, x_j) = 0$ for $1 \leq j \leq m$. Thus, we may assume that $K(x_i, x_i) > 0$ for $1 \leq i \leq m$. We have

$$\sum_{i,j} c_i c_j K'(x_i, x_j) = \sum_{i,j} c_i c_j \frac{K(x_i, x_j)}{\sqrt{K(x_i, x_i)K(x_j, x_j)}}$$

$$= \sum_{i,j} c_i c_j \frac{\langle \Phi(x_i), \Phi(x_j) \rangle}{\| \Phi(x_i) \|_H \| \Phi(x_j) \|_H}$$

$$= \| \sum_i c_i \Phi(x_i) \|_H \geq 0,$$

where $\Phi$ is the feature mapping associated to $K$. 
Example

Let $K$ be the kernel

$$K(u, v) = e^{\frac{u'v}{\sigma^2}},$$

where $\sigma > 0$. Note that $K(u, u) = e^{\frac{\|u\|^2}{\sigma^2}}$ and $K(v, v) = e^{\frac{\|v\|^2}{\sigma^2}}$, hence its normalized kernel is

$$K'(u, v) = \frac{K(u, v)}{\sqrt{K(u, u)} \sqrt{K(v, v)}}$$

$$= \frac{e^{\frac{u'v}{\sigma^2}}}{\sqrt{e^{\frac{\|u\|^2}{2\sigma^2}}} \sqrt{e^{\frac{\|v\|^2}{2\sigma^2}}}}$$

$$= e^{\frac{-\|u-v\|^2}{2\sigma^2}}$$
Example

For a positive constant $\sigma$ a **Gaussian kernel or a radial basis function** is the function $K : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ defined by

$$K(u, v) = e^{-\frac{\|u-v\|^2}{2\sigma^2}}.$$ 

We prove that $K$ is of positive type by showing that $K(x, y) = (\phi(x), \phi(y))$, where $\phi : \mathbb{R}^k \rightarrow \ell^2(\mathbb{R})$. Note that for this example $\phi$ ranges over an infinite-dimensional space.
We have

\[ K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} \]

\[ = e^{-\frac{\|x\|^2 + \|y\|^2 - 2(x, y)}{2\sigma^2}} \]

\[ = e^{-\frac{\|x\|^2}{2\sigma^2}} \cdot e^{-\frac{\|y\|^2}{2\sigma^2}} \cdot e^{\frac{(x, y)}{\sigma^2}} \]
Taking into account that
\[
e^{\frac{(x,y)}{\sigma^2}} = \sum_{j=0}^{\infty} \frac{1}{j!} \frac{(x,y)^j}{\sigma^{2j}}
\]
we can write
\[
\sum_{j=0}^{\infty} \frac{(x,y)^j}{j! \sigma^{2j}} e^{-\frac{\|x\|^2}{2\sigma^2}} \cdot e^{-\frac{\|y\|^2}{2\sigma^2}} = \sum_{j=0}^{\infty} \left( e^{-\frac{\|x\|^2}{2j\sigma^2}} \frac{1}{\sqrt{j!}} \right)^j \left( e^{-\frac{\|y\|^2}{2j\sigma^2}} \frac{1}{\sqrt{j!}} \right)^j (x,y) = (\phi(x), \phi(y)),
\]
where
\[
\phi(x) = \left( \ldots, \frac{e^{-\frac{\|x\|^2}{2j\sigma^2}}}{\sigma j \sqrt{j!}} \left( \frac{1}{\sqrt{j!}} \right)^{\frac{1}{2}} x_1^{n_1} \cdots x_k^{n_k}, \ldots \right).
\]
j varies in \(\mathbb{N}\) and \(n_1 + \cdots + n_k = j\).
Example

For $a, b \geq 0$, a *sigmoid kernel* is defined as

$$K(x, y) = \tanh(ax'y + b)$$

With $a, b \geq 0$ the kernel is of positive type.