The Kernel Trick

Robert M. Haralick

Computer Science, Graduate Center
City University of New York

Outline

SVM Classification

$< (x_1, c_1), \ldots, (x_Z, c_Z) >$	is the training data
$c_1, \ldots, c_Z \in \{-1, 1\}$	specifies the corresponding classes
$\alpha_1, \ldots, \alpha_Z$	are the Langrange multipliers

Assign x to class $+1$ when $w'x > \theta$ Else assign to class -1

$$\begin{aligned} w'x &= \left(\sum_{z \in S} \alpha_z c_z x_z \right)' x \\ &= \sum_{z \in S} \alpha_z c_z (x_z' x) \end{aligned}$$

Dual Form

$$\text{Minimize } L(\alpha_1, \ldots, \alpha_Z) = \sum_{z=1}^{Z} \alpha_z - \frac{1}{2} \sum_{i=1}^{Z} \sum_{j=1}^{Z} \alpha_i \alpha_j c_i c_j (x_i' x_j)$$

subject to

$$\begin{aligned} \alpha_z &\geq 0, \; z = 1, \ldots, Z \\ \sum_{z=1}^{Z} \alpha_z c_z &= 0 \end{aligned}$$

Making Non-Linear Boundaries

Example: Change $x^{2\times 1} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ To $\phi(x)^{5\times 1} = \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_1 x_2 \\ x_1 \\ x_2 \end{pmatrix}$

$$\text{Minimize } L(\alpha_1, \ldots, \alpha_Z) = \sum_{z=1}^{Z} \alpha_z - \frac{1}{2} \sum_{i=1}^{Z} \sum_{j=1}^{Z} \alpha_i \alpha_j c_i c_j (\phi(x_i)' \phi(x_j))$$

subject to

$$\begin{aligned} \alpha_z &\geq 0, \ z = 1, \ldots, Z \\ \sum_{z=1}^{Z} \alpha_z c_z &= 0 \end{aligned}$$

$$w = \sum_{z \in S} \alpha_z c_z \phi(x_z)$$

Making Non-Linear Boundaries

Assign x to class $+1$ when

$$\begin{aligned} w'\phi(x) &> \theta \\ w_1 x_1^2 + w_2 x_2^2 + w_3 x_1 x_2 + w_4 x_1 + w_5 x_2 &> \theta \end{aligned}$$

- The linear problem in the 5-dimensional space is the *same* as the quadratic problem in the 2-dimensional space.
- The kernel trick avoids the explicit mapping that is needed to get linear learning algorithms to learn a nonlinear function or decision boundary.

Making Non-Linear Boundaries

$$\text{Minimize } L(\alpha_1, \ldots, \alpha_Z) = \sum_{z=1}^{Z} \alpha_z - \frac{1}{2} \sum_{i=1}^{Z} \sum_{j=1}^{Z} \alpha_i \alpha_j c_i c_j (\phi(x_i)' \phi(x_j))$$

Set $K(x, y) = \phi(x)' \phi(y)$

$$\begin{aligned}
\text{Minimize } L(\alpha_1, \ldots, \alpha_Z) &= \sum_{z=1}^{Z} \alpha_z - \frac{1}{2} \sum_{i=1}^{Z} \sum_{j=1}^{Z} \alpha_i \alpha_j c_i c_j K(x_i, x_j) \\
w' \phi(x) &= \sum_{z \in S} \alpha_z c_z \phi(x_z)' \phi(x) \\
&= \sum_{z \in S} \alpha_z c_z K(x_z, x)
\end{aligned}$$

The Kernel Trick

What happens when $K(x, y) = (x'y)^2$?
Example:

$$
\begin{aligned}
x &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
y &= \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\
x'y &= x_1 y_1 + x_2 y_2 \\
(x'y)^2 &= x_1^2 y_1^2 + 2 x_1 y_1 x_2 y_2 + x_2^2 y_2^2 \\
&= \begin{pmatrix} x_1^2 \\ \sqrt{2} x_1 x_2 \\ x_2^2 \end{pmatrix}' \begin{pmatrix} y_1^2 \\ \sqrt{2} y_1 y_2 \\ y_2^2 \end{pmatrix} \\
&= \phi(x)' \phi(y)
\end{aligned}
$$

The components of the weight vector can compensate for coefficients that are not 1.

Two-Class Data Set

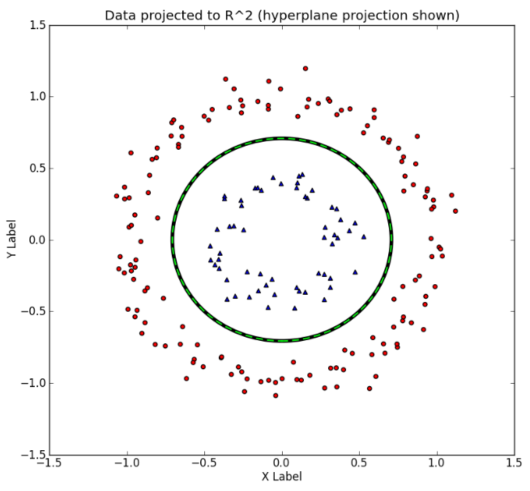

Figure: Raw Data in 2D is not separable with hyperplane.

Two-Class Data Set Transformed

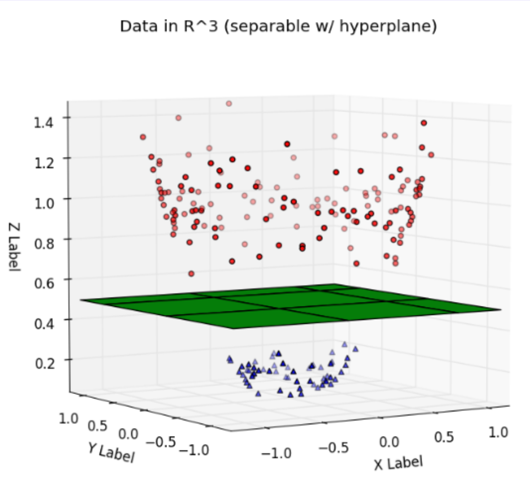

Figure: Transformed: $(x_1, x_2) \mapsto (x_1^2, \sqrt{2}x_1x_2, x_2^2)$

Brute Force: Quadratic Case

If x has N components, to make it quadratic requires an additional $N(N-1)/2$ components

- $x^{N\times 1}$
- $\phi(x)^{(N+N*(N-1)/2)\times 1}$
- $w^{(N+N*(N-1)/2)\times 1} = \sum_{z\in S} \alpha_z c_z \phi(x_z)$
- $w'\phi(x) = \sum_{z\in S} \alpha_z c_z \phi(x_z)'\phi(x)$
- Takes $N + N * (N-1)/2$ operations for $x \mapsto \phi(x)$
- Takes $N + N * (N-1)/2$ operations for $w'\phi(x)$

The Kernel Trick

- $x^{N \times 1}$
- $x'y$ takes N operations
- $(x'y)^2$ takes $N + 1$ operations
- $\sum_{z \in S} \alpha_z c_z K(x_z, x)$
- $K(x_z, x) = (x_z' x)^2$ takes $N + 1$ operations
- Compute $w'x = |S|(N + 1)$ operations
- Brute Force $N + N * (N - 1)/2$ operations

The Kernel Trick

What happens when $K(x, y) = (x'y)^3$?
Example:

$$\begin{aligned}
x'y &= x_1 y_1 + x_2 y_2 \\
(x'y)^3 &= (x_1 y_1)^3 + 3(x_1 y_1)^2 x_2 y_2 + 3(x_1 y_1)(x_2 y_2)^2 + (x_2 y_2)^3 \\
&= \begin{pmatrix} x_1^3 \\ \sqrt{3} x_1^2 x_2 \\ \sqrt{3} x_1 x_2^2 \\ x_2^3 \end{pmatrix}' \begin{pmatrix} y_1^3 \\ \sqrt{3} y_1^2 y_2 \\ \sqrt{3} y_1 y_2^2 \\ y_2^3 \end{pmatrix}
\end{aligned}$$

The Kernel Trick

What kind of functions $K(x, y)$ can be represented in the form

$$K(x, y) = \phi(x)'\phi(y)$$

Kernel

Definition

A Kernel is a function $K : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}$, such that for all $x, y \in \mathbb{R}^M$, there is a feature function $\phi : \mathbb{R}^M \rightarrow \mathbb{R}^N$ such that

$$K(x, y) = \phi(x)'\phi(y)$$

Proposition

Kernels are symmetric functions.

Kernel

Let $x = (x_1, \ldots, x_M)'$ and $y = (y_1, \ldots, y_M)'$

$$
\begin{aligned}
K(x, y) &= (x'y)^2 \\
&= \left(\sum_{m=1}^{M} x_m y_m \right)^2 \\
&= \sum_{i=1}^{M} x_i y_i \sum_{j=1}^{M} x_j y_j \\
&= \sum_{i=1}^{M} \sum_{j=1}^{M} x_i y_i x_j y_j \\
&= \sum_{i=1}^{M} \sum_{j=1}^{M} (x_i x_j)(y_i y_j) \\
&= \left((x_i x_j)|_{(i,j)=(1,1)}^{(M,M)} \right)' \left((y_i y_j)|_{(i,j)=(1,1)}^{(M,M)} \right) \\
\phi(x) &= \left((x_i x_j)|_{(i,j)=(1,1)}^{(M,M)} \right)
\end{aligned}
$$

Kernel

Let $x = (x_1, \ldots, x_M)'$ and $y = (y_1, \ldots, y_M)'$

$$
\begin{aligned}
K(x, y) &= (x'y + 1)^2 \\
&= \left(\sum_{i=1}^{M} x_i y_i + 1 \right) \left(\sum_{j=1}^{M} x_j y_j + 1 \right) \\
&= \sum_{i=1}^{M} \sum_{j=1}^{M} x_i x_j y_i y_j + 2 \sum_{i=1}^{M} x_i y_i + 1 \\
&= \sum_{i=1}^{M} \sum_{j=1}^{M} (x_i x_j)(y_i y_j) + \sum_{i=1}^{M} (\sqrt{2} x_i)(\sqrt{2} y_i) + (1)(1)
\end{aligned}
$$

Common Kernels

$$x, y \in \mathbb{R}^N$$

Polynomial	$K(x,y)$	$=$	$(x'y+1)^q$	$N+1+2$ (log and exponentiate
Radial	$K(x,y)$	$=$	$\exp\left(-\frac{\lVert x-y\rVert^2}{\sigma^2}\right)$	$N+1+1$ (exponentiate)
Sigmoid	$K(x,y)$	$=$	$\tanh(\beta x'y+\gamma)$	$N+1+1$ (hyperbolic tangent)

$K(x,y)$ measures the similarity between x and y

Kernels and Feature Spaces

Given a function $K : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$, how can we tell that there exists a feature function ϕ such that

$$K(x, y) = \phi(x)'\phi(y)$$

thereby making K a kernel?

Mercer's Kernel Characterization Theorem

Let $K : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ be a symmetric function. There exists $\phi : \mathbb{R}^M \to \mathbb{R}^N$ such that

$$K(x, y) = \phi(x)'\phi(y)$$

if and only if

$$\int_{x \in \mathbb{R}^M} \int_{y \in \mathbb{R}^M} K(x, y) g(x) g(y) dx dy \geq 0$$

for every function $g : \mathbb{R}^M \to \mathbb{R}$ satisfying

$$\int_{x \in \mathbb{R}^M} g(x)^2 dx < \infty$$

Mercer's Kernel Characterization Theorem

Let $K : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ be a symmetric function. There exists $\phi : \mathbb{R}^M \to \mathbb{R}^N$ such that

$$K(x, y) = \phi(x)'\phi(y)$$

if and only if for every L and every set $\{x_1, \ldots, x_L\}$ of L points from R^M, the matrix $\mathbb{K}$,

$$\mathbb{K} = \begin{pmatrix} K(x_1, x_1) & K(x_1, x_2) & \ldots, & K(x_1, x_L) \\ K(x_2, x_1) & K(x_2, x_2) & \ldots, & K(x_2, x_L) \\ & \vdots & \vdots & \\ K(x_L, x_1) & K(x_L, x_2) & \ldots, & K(x_L, x_L) \end{pmatrix}$$

is positive semidefinite.

Symmetric Positive Semi-definite

Definition

A symmetric matrix $B^{N \times N}$ is Positive Semi-definite if and only if for every $x \in \mathbb{R}^N$,

$$x'Bx \geq 0$$

Theorem

A symmetric matrix B is positive semi-definite if and only if all its eigenvalues are non-negative.

Theorem

If B is a symmetric positive semi-definite matrix, then there exists a matrix A satisfying $B = A'A$.

Positive Definite Kernel

Definition

A symmetric function $K : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ is Positive Definite (pd) if and only if for any $x_1, \ldots, x_L \in \mathbb{R}^N$ and any $c_1, \ldots, c_L \in R$

$$\sum_{i=1}^{L} \sum_{j=1}^{L} c_i K(x_i, x_j) c_j \geq 0$$

An alternative way of writing $\sum_{i=1}^{L} \sum_{j=1}^{L} c_i K(x_i, x_j) c_j$ is $c' \mathbb{K} c$ where $c' = (c_1, \ldots, c_L)$ and $\mathbb{K}_{ij} = K(x_i, x_j)$

A Positive Definite Kernel

Proposition

Let $K : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *be defined by* $K(x, y) = x'y$*. Then* K *is a positive definite kernel.*

Proof.

Let $x_1, \ldots, x_K \in \mathbb{R}^N$*. Define*

$$\begin{aligned} A &= \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_K' \end{pmatrix} \\ \mathbb{K} &= AA' \end{aligned}$$

Let $c \in \mathbb{R}^K$*.*

$$\begin{aligned} c'\mathbb{K}c &= c'AA'c \\ &= (A'c)'(A'c) \geq 0 \end{aligned}$$

Positive Definite Kernel Properties

Proposition

Suppose $K_i : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$, $i = 1, 2, \ldots,$ are positive definite (pd) kernels then

- $K(x, y) = \sum_{j=1}^{J} a_j K_j(x, y)$ *is a pd kernel when* $a_j \geq 0$
- $K(x, y) = \prod_{j=1}^{J} K_j(x, y)$ *is a pd kernel*
- $K(x, y) = \lim_{j \to \infty} K_j(x, y)$ *is a pd kernel*
- $K(x, y) = q(K_1(x, y))$ *is a pd kernel where* q *is any polynomial with non-negative coefficients*
- $K(x, y) = f(x)K_1(x, y)f(y)$ *is a pd kernel for any* $f : \mathbb{R}^N \to \mathbb{R}$
- $K(x, y) = f(x)f(y)$ *is a pd kernel for any* $f : \mathbb{R}^N \to \mathbb{R}$
- $K(x, y) = K_1(\phi(x), \phi(y))$
- $K(x, y) = x'Bz$, *where* B *is symmetric positive semidefinite matrix*

Products of pd kernels are pd kernels

Proposition

Let $\psi_1, \psi_2 : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *be positive definite kernels. Define* $\psi : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *by* $\psi(x, y) = \psi_1(x, y)\psi_2(x, y)$ *Then* ψ *is a positive definite kernel.*

Proof.

Let $x_1, \ldots, x_M \in R^N$ *and let* $c_1, \ldots, c_M \in \mathbb{R}$. *Then*

$$\sum_{i=1}^{M}\sum_{j=1}^{M} c_i \psi(x_i, x_j) c_j = \sum_{i=1}^{M}\sum_{j=1}^{M} c_i \psi_1(x_i, x_j)\psi_2(x_i, x_j) c_j$$

Since ψ_1 *and* ψ_2 *are both positive definite, there exists matrices* $A^{M\times M}$ *and* $B^{M\times M}$ *such that*

$$\begin{aligned} \psi_1(x_i, x_j) &= \left(A'A\right)_{ij} = \sum_{k=1}^{M} a_{ki} a_{kj} \\ \psi_2(x_i, x_j) &= \left(B'B\right)_{ij} = \sum_{k=1}^{K} b_{ki} b_{kj} \end{aligned}$$

Products of pd kernels are pd kernels

Proof.

Hence, $\psi_1(x_i, x_j) = \sum_{k=1}^{M} a_{ki} a_{kj}$ and $\psi_2(x_i, x_j) = \sum_{l=1}^{M} b_{li} b_{lj}$. Therefore,

$$\begin{aligned}
\sum_{i=1}^{M}\sum_{j=1}^{M} c_i \psi_1(x_i, x_j)\psi_2(x_i, x_j) c_j &= \sum_{i=1}^{M}\sum_{j=1}^{M} c_i \sum_{k=1}^{M} a_{ki} a_{kj} \sum_{l=1}^{M} b_{li} b_{lj} \\
&= \sum_{k=1}^{M}\sum_{l=1}^{M} \left(\sum_{i=1}^{M} c_i a_{ki} b_{li}\right)\left(\sum_{j=1}^{M} c_j a_{kj} b_{lj}\right) \\
&= \sum_{k=1}^{M}\sum_{l=1}^{M} \left(\sum_{i=1}^{M} c_i a_{ki} b_{li}\right)^2 \geq 0
\end{aligned}$$

$\square$

Exponential Kernel

Proposition

Define $q_j(x) = \sum_{n=0}^{j} \frac{x^n}{n!}$. *Suppose* $K : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *is a positive definite kernel, then,*

$$\exp\left(K(x,y)\right) = lim_{j\to\infty} q_j(K(x,y))$$

is a positive definite kernel.

Proof.

$q_j(x)$ *is a polynomial with all positive coefficients. Since* K *is a positive definite kernel,* $q_j(K(x,y))$ *is a positive definite kernel. Limits of positive definite kernels are positive definite kernels. Therefore* $\exp\left(K(x,y)\right)$ *is a positive definite kernel.* □

Positive Definite Kernel Property

Proposition

Let $\psi : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be a symmetric positive definite kernel. Let $f : \mathbb{R}^N \to \mathbb{R}$ be any function. Then $K(x, y) = f(x)\psi(x, y)f(y)$ is a symmetric positive definite kernel.

Proof.

The symmetry is immediate because ψ is symmetric and multiplication is commutative. Let $x_1, \ldots, x_M \in \mathbb{R}^N$ and $c_1, \ldots, c_M \in \mathbb{R}$. Then

$$\begin{aligned} \sum_{i=1}^{M}\sum_{j=1}^{M} c_i K(x_i, x_j) c_j &= \sum_{i=1}^{M}\sum_{j=1}^{M} c_i f(x_i)\psi(x_i, x_j) f(x_j c_j) \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M} (c_i f(x_i))\psi(x_i, x_j)(f(x_j)c_j) \end{aligned}$$

Let $d_i = c_i f(x_i)$. Then, since ψ is positive definite,

$$\sum_{i=1}^{M}\sum_{j=1}^{M} c_i K(x_i, x_j) c_j = \sum_{i=1}^{M}\sum_{j=1}^{M} d_i \psi(x_i, x_j) d_j \geq 0$$

Positive Definite Kernel

Proposition

Let $K : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *be a positive definite kernel. Then* $K(x,x)K(y,y) \geq K(x,y)^2$.

Proof.

Let $x, y \in R^N$. *Then the matrix* $\begin{pmatrix} K(x,x) & K(x,y) \\ K(y,x) & K(y,y) \end{pmatrix}$ *is positive semidefinite. Hence the determinant must be non-negative so that* $K(x,x)K(y,y) - K(x,y)^2 \geq 0$. *Therefore* $K(x,x)K(y,y) \geq K(x,y)^2$. $\square$

Example

Proposition

Let $K : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *be defined by*

$$K(x, y) = \exp\left(-\frac{(x-y)'(x-y)}{\sigma^2}\right)$$

Then K *is a positive definite kernel.*

Proof.

$x'y$ *is a pd kernel.*
$\frac{2x'y}{\sigma^2}$ *is a pd kernel.*
$\exp\left(\frac{2x'y}{\sigma^2}\right)$ *is a pd kernel.*
$\exp\left(-\frac{x'x}{\sigma^2}\right)\exp\left(\frac{2x'y}{\sigma^2}\right)\exp\left(-\frac{y'y}{\sigma^2}\right)$ *is a pd kernel.*
$\exp(-\frac{(x-y)'(x-y)}{\sigma^2})$ *is a pd kernel.* $\square$

Positive Definite Kernel Properties

Proposition

Let $K : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *be a positive definite kernel. Then*

$$K_1(x, y) = \frac{K(x, y)}{\sqrt{K(x, x)K(y, y)}}$$

is a positive definite kernel.

Proof.

$$\begin{aligned} K_1(x, y) &= \frac{K(x, y)}{\sqrt{K(x, x)K(y, y)}} \\ &= \frac{1}{\sqrt{K(x, x)}} K(x, y) \frac{1}{\sqrt{K(y, y)}} \end{aligned}$$

$\square$

Positive Definite Kernel

Proposition

Let $h : \mathbb{R}^N \to \mathbb{R}$ be any function satisfying $h(x) \geq 0$ with minimum at 0. Then,

$$K(x, y) = h(x + y) - h(x - y)$$

is a positive definite kernel.

Symmetric Negative Definite

Definition

A symmetric function $\psi : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is Negative Definite if and only if for every $x_1, \ldots, x_M \in \mathbb{R}^N$ and $c_1, \ldots, c_M \in \mathbb{R}$ satisfying $\sum_{m=1}^{M} c_m = 0$

$$\sum_{i=1}^{M} \sum_{j=1}^{M} c_i \psi(x_i, x_j) c_j \leq 0$$

Proposition

Let $\Psi^{M \times M}$ be defined by $\Psi_{ij} = \psi(x_i, x_j)$. If Ψ is negative semi-definite Then ψ is negative definite.

Properties of Kernels

Proposition

Let $K : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be a symmetric function. If K is a positive definite kernel then $-K$ is a negative definite kernel.

Proof.

Let $x_1, \ldots, x_K \in \mathbb{R}^N$ and $c \in \mathbb{R}^K$ Suppose K is a positive definite kernel. then $c'\mathbb{K}c \geq 0$. Then $c'(-\mathbb{K})c = -c'\mathbb{K}c \leq 0$. Since this is true for all c, it is certainly true for c satisfying $1'c = 0$ and this implies that $-K$ is a negative definite kernel. $\square$

A Negative Definite Kernel

Proposition

Let $\psi : \mathbb{R}^N \times \mathbb{R}^N$ be defined by $\psi(x, y) = (x - y)'(x - y)$. Then ψ is a negative definite kernel.

Proof.

Let $x_1, \ldots, x_K \in \mathbb{R}^N$ and $c_1, \ldots, c_K \in \mathbb{R}$ satisfying $\sum_{k=1}^K c_k = 0$. Then,

$$\begin{aligned}
\sum_{i=1}^K \sum_{j=1}^K c_i (x_i - x_j)'(x_i - x_j) c_j &= \sum_{i=1}^K c_i x_i' x_i \sum_{j=1}^K c_j - 2 \sum_{i=1}^K \sum_{j=1}^K c_i x_i' x_j c_j + \\
&\quad \sum_{i=1}^K c_i \sum_{j=1}^K x_j' x_j c_j \\
&= -2 \sum_{i=1}^K \sum_{j=1}^K c_i x_i' x_j c_j \leq 0
\end{aligned}$$

$\square$

Negative Definite Kernel Properties

Proposition

Let $K_m : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$, $m = 1, 2, \ldots,$ be symmetric and negative definite. Then J_1, J_2, J_3, J_4 and J_5 defined by

$$
\begin{aligned}
J_1(x, y) &= \sum_{m=1}^{M} a_m K_m(x, y), \;\; a_m \geq 0 \\
J_2(x, y) &= \lim_{m \to \infty} K_m(x, y) \\
J_3(x, y) &= log(1 + K(x, y)) \\
J_4(x, y) &= K(x, y)^{\alpha}, \;\; 0 < \alpha < 1 \\
J_5(x, y) &= f(x) + f(y), \;\; \textit{for any } f : \mathbb{R}^N \to \mathbb{R}
\end{aligned}
$$

are negative definite kernels.

Negative Definite Kernel

Proposition

Let $\psi : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *be a symmetric negative definite kernel. Then* $\psi(x,x) + \psi(y,y) \leq 2\psi(x,y)$

Proof.

Let $x, y \in \mathbb{R}$ *and* $c_1, c_2 \in \mathbb{R}$ *satisfying* $c_1 + c_2 = 0$. *Since* ψ *is negative definite,*

$$\begin{aligned}
0 &\geq c_1\psi(x,x)c_1 + c_1\psi(x,y)c_2 + c_2\psi(y,x)c_1 + c_2\psi(y,y)c_2 \\
&\geq c_1\psi(x,x)c_1 + c_1\psi(x,y)(-c_1) + (-c_1)\psi(y,x)c_1 + (-c_1)\psi(y,y)(-c_1) \\
&\geq \psi(x,x) - 2\psi(x,y) + \psi(y,y) \\
2\psi(x,y) &\geq \psi(x,x) + \psi(y,y)
\end{aligned}$$

$\square$

Positive and Negative Definite Symmetric Functions

Theorem

(Shoenberg,1938)
Let $\psi : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be symmetric. Then $\psi(x, y)$ is negative definite if and only if $\exp(-t\psi(x, y)$ is positive definite for all $t > 0$.

Proposition

Let $K : \mathbb{R}^n \times \mathbb{R}^N \to \mathbb{R}$ be a positive definite kernel satisfying $K(x, y) \geq 0$, $x, y \in \mathbb{R}^N$. Then

$$-logK(x, y)$$

is negative definite if and only if

$$K(x, y)^t$$

is positive definite for all $t > 0$

Positive and Negative Definite Symmetric Functions

Proposition

Let $\psi : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *be a symmetric function. Let* $z \in \mathbb{R}^N$. *Define*

$$\phi(x, y) = -\psi(x, y) + \psi(x, z) + \psi(z, y) - \psi(z, z)$$

Then ψ *is negative definite if and only if* ϕ *is positive definite.*

Proof.

Suppose ψ *is negative definite. Let* $x_k \in \mathbb{R}^N, k = 0, \ldots, K$ *and* $c_k \in \mathbb{R}, k = 1, \ldots, K$ *with* $c_0 = -\sum_{k=1}^{K} c_k$. *Then since* ψ *is negative definite,*

$$0 \geq \sum_{i=0}^{K} \sum_{j=0}^{K} c_i \psi(x_i, x_j) c_j$$

$$\begin{aligned} 0 \geq \sum_{i=0}^{K} \sum_{j=0}^{K} c_i \psi(x_i, x_j) c_j \quad &= \quad \sum_{i=1}^{K} \sum_{j=1}^{K} c_i \psi(x_i, x_j) c_j + c_0 \sum_{j=1}^{K} \psi(z, x_j) c_j + \\ & \quad c_0 \sum_{i=1}^{K} c_i \psi(z, x_j) + c_0^2 \psi(z, z) \end{aligned}$$

Positive and Negative

Proof.

$$\begin{aligned}
\sum_{i=0}^{K}\sum_{j=0}^{K} c_i\psi(x_i,x_j)c_j &= \sum_{i=1}^{K}\sum_{j=1}^{K} c_i\psi(x_i,x_j)c_j - \sum_{i=1}^{K} c_i \sum_{j=1}^{K}\psi(z,x_j)c_j - \\
&\quad \sum_{i=1}^{K}\sum_{j=1}^{K} c_j c_i\psi(x_i,z) + \sum_{i=1}^{K} c_i \sum_{j=1}^{K} c_j\psi(z,z) \\
&= \sum_{i=1}^{K}\sum_{j=1}^{K} c_i(\psi(x_i,x_j)-\psi(x_i,z)-\psi(z,x_j)+\psi(z,z))c_j \\
0 &\geq -\sum_{i=1}^{K}\sum_{j=1}^{K} c_i\phi(x_i,x_j)c_j \\
0 &\leq \sum_{i=1}^{K}\sum_{j=1}^{K} c_i\phi(x_i,x_j)c_j
\end{aligned}$$

This make ϕ positive definite. $\square$

Positive and Negative

Proof.

Suppose ϕ is positive definite. Let $c_1, \ldots c_K \in \mathbb{R}$ satisfy $\sum_{k=1}^{K} c_k = 0$.Then

$$\sum_{i=1}^{K}\sum_{j=1}^{K} c_i \phi(x_i, x_j) c_j \geq 0$$

Since $\phi(x, y) = -\psi(x, y) + \psi(x, z) + \psi(z, y) - \psi(z, z)$

$$\begin{aligned}
0 &\leq \sum_{i=1}^{K}\sum_{j=1}^{K} c_i(-\psi(x_i, x_j) + \psi(x_i, z) + \psi(z, x_j) - \psi(z, z))c_j \\
&\leq -\sum_{i=1}^{K}\sum_{j=1}^{K} c_i\psi(x_i, x_j)c_j + \sum_{i=1}^{K} c_i\psi(x_i, z)\sum_{j=1}^{K} c_j + \\
&\quad \sum_{i=1}^{K} c_i \sum_{j=1}^{K} \psi(z, x_j)c_j - \psi(z, z)\sum_{i=1}^{K} c_i \sum_{j=1}^{K} c_j = -\sum_{i=1}^{K}\sum_{j=1}^{K} c_i\psi(x_i, x_j)c_j
\end{aligned}$$

Therefore ψ is negative definite. □

Negative Definite From Positive Definite

Proposition

Let $K : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be a positive definite symmetric function. Define $\phi(x,y) = K(x,x) + K(y,y) - 2K(x,y)$. Then ϕ is negative definite.

Proof.

Let $x_1, \ldots, x_M \in \mathbb{R}^N$ and $c_1, \ldots, c_M \in \mathbb{R}$ satisfying $\sum_{m=1}^M c_m = 0$. Then,

$$\begin{aligned}
\sum_{i=1}^M \sum_{j=1}^M c_i \phi(x_i, x_j) c_j &= \sum_{i=1}^M \sum_{j=1}^M c_i (K(x_i, x_i) + K(x_j, x_j) - 2K(x_i, x_j)) c_j \\
&= \sum_{i=1}^M c_i K(x_i, x_i) \sum_{j=1}^M c_j + \sum_{j=1}^M c_j K(x_j, x_j) \sum_{i=1}^M c_i + \\
&\quad -2 \sum_{i=1}^M \sum_{j=1}^M c_i K(x_i, x_j) c_j \leq 0
\end{aligned}$$

$\square$

Kernel Properties

Proposition

Let $K : \mathbb{R}^n \times \mathbb{R}^N \to \mathbb{R}$ *be a positive definite kernel satisfying* $K(x,y) \geq 0$, $x, y \in \mathbb{R}^N$. *Then*

$$-logK(x,y)$$

is negative definite if and only if

$$K(x,y)^t$$

is positive definite for all $t > 0$

Kernel Properties

Proposition

Let $K : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *be a negative definite kernel that satisfies*

$$K(x, y) > 0$$

Then,

$$\frac{1}{K(x, y)}$$

is a positive definite kernel.

Other Positive Definite Kernels

- Laplacian: $K(x,y) = \exp\left(-\frac{\|x-y\|}{\sigma}\right)$
- ANOVA: $K(x,y) = \sum_{n=1}^{N} \exp(-\sigma(x^n - y^n)^2)^d$
- Rational Quadratic: $K(x,y) = 1 - \frac{\|x-y\|^2}{\|x-y\|^2+c}$
- Multiquadric: $K(x,y) = \sqrt{\|x-y\|^2 + c^2}$
- Inverse Multiquadric: $K(x,y) = \frac{1}{\sqrt{\|x-y\|^2+c^2}}$
- Circular: $K(x,y) = \frac{2}{\pi}\arccos(-\frac{\|x-y\|}{\sigma}) - \frac{2}{\pi}\frac{\|x-y\|}{\sigma}\sqrt{1-\left(\frac{\|x-y\|}{\sigma}\right)^2}$ if $\|x-y\| < \sigma$, zero otherwise
- Spherical: $K(x,y) = 1 - \frac{3}{2}\frac{\|x-y\|}{\sigma} + \frac{1}{2}\left(\frac{\|x-y\|}{\sigma}\right)^3$ if $\| x - z \| < \sigma$, zero otherwise
- Wave Kernel: $K(x,y) = \frac{\theta}{\|x-y\|}\sin\frac{\|x-y\|}{\theta}$
- Spline:
  $K(x,y) = \prod_{n=1}^{N}(1 + x_n y_n + x_n y_n \min(x_n, y_n) - \frac{x_n+y_n}{2}\min(x_n,y_n)^2 + \frac{\min(x_n,y_n)^3}{3})$
- Cauchy: $K(x,y) = \frac{1}{1+\frac{\|x-y\|^2}{\sigma^2}}$
- Chi-Square: $K(x,y) = 1 - \sum_{n=1}^{N}\frac{(x_n-y_n)^2}{\frac{1}{2}(x_n+y_n)}$
- Histogram: $K(x,y) = \sum_{n=1}^{N}\min(x_n,y_n)$, where $x = (x_1, \ldots, x_N)$ and $y = (y_1, \ldots, y_N)$