Positive Definite Kernel Functions on Fuzzy Sets

FUZZ 2014

Jorge Guevara Díaz[1] Roberto Hirata Jr.[1] Stéphane Canu[2]

[1]Institute of Mathematics and Statistics
University of Sao Paulo-Brazil

[2]Institut National de Sciences Appliquees
University of Normandy-France

July 08, 2014

Table of Contents

Table of Contents

Data representation

Machine learning on vector spaces (feature space)

Data representation

- Machine learning on vector spaces (feature space)
- Data representation: graphs, sets, distributions, logic terms, *fuzzy sets*

Data representation

- Machine learning on vector spaces (feature space)
- Data representation: graphs, sets, distributions, logic terms, *fuzzy sets*
- Kernel methods: Data Representation $\rightarrow$ RKHS

Table of Contents

Fuzzy Set

Let Ω be a nonempty set. A fuzzy set on Ω, is a set $X \subseteq \Omega$ with membership function

$$\mu_X : \Omega \rightarrow [0,1] \quad (1)$$
$$x \mapsto \mu_X(x). \quad (2)$$

Definition (Support of a fuzzy set)

The support of a fuzzy set is the set

$$X_{>0} = \{x \in \Omega | \mu_X(x) > 0\}.$$

T-Norm

A triangular norm or T-norm is the function $T : [0,1]^2 \rightarrow [0,1]$, such that, for all $x, y, z \in [0,1]$ satisfies:

- T1 commutativity: $T(x,y) = T(y,x)$;
- T2 associativity: $T(x, T(y,z)) = T(T(x,y), z)$;
- T3 monotonicity: $y \leq z \Rightarrow T(x,y) \leq T(x,z)$;
- T4 boundary condition $T(x,1) = x$.

T-Norm

A triangular norm or T-norm is the function $T : [0,1]^2 \to [0,1]$, such that, for all $x, y, z \in [0,1]$ satisfies:

- T1 commutativity: $T(x,y) = T(y,x)$;
- T2 associativity: $T(x, T(y,z)) = T(T(x,y), z)$;
- T3 monotonicity: $y \leq z \Rightarrow T(x,y) \leq T(x,z)$;
- T4 boundary condition $T(x,1) = x$.

a multiple-valued extension

Using $n \in \mathbb{N}, n \geq 2$ and associativity, a multiple-valued extension $T_n : [0,1]^n \to [0,1]$ of a T-norm T is given by $T_2 = T$ and

$$T_n(x_1, x_2, \ldots, x_n) = T(x_1, T_{n-1}(x_2, x_3, \ldots, x_n)). \tag{3}$$

We will use T to denote T or T_n.

Semi-ring of sets

A semi-ring of sets, $\mathcal{S}$ on Ω, is a subset of the power set $\mathcal{P}(\Omega)$, that is, a set of sets satisfying:

1. $\phi \in \mathcal{S}$, ϕ denotes the empty set;
2. $A, B \in \mathcal{S}, \implies A \cap B \in \mathcal{S}$;
3. for all $A, A_1 \in \mathcal{S}$ and $A_1 \subseteq A$, there exists a sequence of pairwise disjoint sets $A_2, A_3, \ldots A_N \subseteq \mathcal{S}$, such

$$A = \bigcup_{i=1}^{N} A_i.$$

Semi-ring of sets

A semi-ring of sets, $\mathcal{S}$ on Ω, is a subset of the power set $\mathcal{P}(\Omega)$, that is, a set of sets satisfying:

1. $\phi \in \mathcal{S}$, ϕ denotes the empty set;
2. $A, B \in \mathcal{S}, \implies A \cap B \in \mathcal{S}$;
3. for all $A, A_1 \in \mathcal{S}$ and $A_1 \subseteq A$, there exists a sequence of pairwise disjoint sets $A_2, A_3, \ldots A_N \subseteq \mathcal{S}$, such

$$A = \bigcup_{i=1}^{N} A_i.$$

Finite decomposition

Condition 3 is called *finite decomposition of A*.

Measure

Definition (Measure)

Let $\mathcal{S}$ be a semi-ring and let $\rho : \mathcal{S} \to [0, \infty]$ be a pre-measure, i.e., ρ satisfy:

1. $\rho(\phi) = 0$;
2. for a finite decomposition of $A \in \mathcal{S}$, $\rho(A) = \sum_{i=1}^{N} \rho(A_i)$;

by *Carathéodory's extension theorem*, ρ is a measure on $\sigma(\mathcal{S})$, where $\sigma(S)$ is the smallest σ-algebra containing S.

Reproducing Kernel Hilbert Spaces

Definition (Reproducing kernel)

A function

$$\begin{aligned} k : E \times E &\rightarrow \mathbb{R} \\ (x, y) &\mapsto k(x, t) \end{aligned} \tag{4}$$

is called a *reproducing kernel* of the Hilbert space $\mathcal{H}$ if and only if:

1. $\forall x \in E,\ k(., x) \in \mathcal{H}$
2. $\forall x \in E,\ \forall f \in \mathcal{H}\ \ \langle f, k(., x)\rangle_{\mathcal{H}} = f(x)$

Reproducing Kernel Hilbert Spaces

Definition (Reproducing kernel)

A function

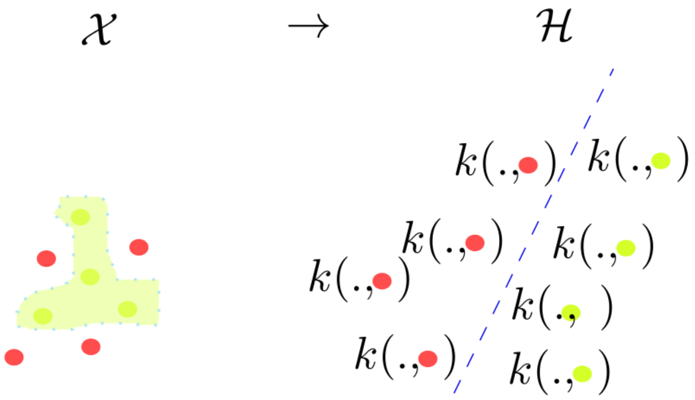
$$\begin{aligned} k : E \times E &\rightarrow \mathbb{R} \\ (x, y) &\mapsto k(x, t) \end{aligned} \tag{4}$$

is called a *reproducing kernel* of the Hilbert space $\mathcal{H}$ if and only if:

1. $\forall x \in E, \; k(., x) \in \mathcal{H}$
2. $\forall x \in E, \; \forall f \in \mathcal{H} \;\; \langle f, k(., x) \rangle_{\mathcal{H}} = f(x)$

Reproducing property

$$\forall (x, y) \in E \times E, \; k(x, y) = \langle k(., x), k(., y) \rangle_{\mathcal{H}} \tag{5}$$



Why kernels?

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

Reproducing Kernel Hilbert Spaces

Definition (Real RKHS)

A Hilbert Space of real valued functions on E, denoted by $\mathcal{H}$, with reproducing kernel is called a real Reproducing Kernel Hilbert Space or real RKHS.

Reproducing Kernel Hilbert Spaces

Definition (Real RKHS)

A Hilbert Space of real valued functions on E, denoted by $\mathcal{H}$, with reproducing kernel is called a real Reproducing Kernel Hilbert Space or real RKHS.

Characterization

All the evaluation functionals are continuous on $\mathcal{H}$. :

$$
\begin{aligned}
e_x : \mathcal{H} &\rightarrow \mathbb{R} && (6)\\
f &\mapsto e_x(f) = f(x) && (7)
\end{aligned}
$$

Reproducing Kernel Hilbert Spaces

Definition (Real RKHS)

A Hilbert Space of real valued functions on E, denoted by $\mathcal{H}$, with reproducing kernel is called a real Reproducing Kernel Hilbert Space or real RKHS.

Characterization

All the evaluation functionals are continuous on $\mathcal{H}$. :

$$e_x : \mathcal{H} \rightarrow \mathbb{R} \quad (6)$$

$$f \mapsto e_x(f) = f(x) \quad (7)$$

A sequence converging in the norm also converges pointwise

By Riez representation theorem and the reproducing property it follows that $\forall x \in E$, $\forall f \in \mathcal{H}$:

$$e_x(f) = f(x) = \langle f, k(., x) \rangle_{\mathcal{H}} \leq \|f\|_{\mathcal{H}} \, \|k\|_{\mathcal{H}}$$

Positive Definite Kernel

Lema

Any reproducing kernel $k : E \times E \to \mathbb{R}$ is a symmetric positive definite function, that is, it satisfies:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(x_i, x_j) \geq 0 \tag{8}$$

$\forall N \in \mathbb{N}$, $\forall c_i, c_j \in \mathbb{R}$ *and* $k(x, y) = k(y, x)$, $\forall x, y \in E$. *The converse is true.*

Positive Definite Kernel

Lema

Any reproducing kernel $k : E \times E \to \mathbb{R}$ is a symmetric positive definite function, that is, it satisfies:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k(x_i, x_j) \geq 0 \tag{8}$$

$\forall N \in \mathbb{N}$, $\forall c_i, c_j \in \mathbb{R}$ *and* $k(x, y) = k(y, x)$, $\forall x, y \in E$. *The converse is true.*

Consequently

Kernels $k(.,.)$ are reproducing kernels of some RKHS. The space spanned by $k(x,.)$ generates a RKHS or a Hilbert space with reproducing kernel k.

Positive Definite Kernel

If k is a reproducing kernel, then

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(x_i, x_j) &= \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \langle k(., x_i), k(., x_j)\rangle_{\mathcal{H}} \\
&= \langle \sum_{i=1}^{N} c_i k(., x_i), \sum_{j=1}^{N} c_j k(., x_j)\rangle_{\mathcal{H}} \\
&= \| \sum_{i=1}^{N} c_i k(., x_i)\|_{\mathcal{H}}^2 \\
&\geq 0
\end{aligned}
$$

Positive Definite Kernel

If k is a reproducing kernel, then

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(x_i, x_j) &= \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \langle k(., x_i), k(., x_j)\rangle_{\mathcal{H}} \\
&= \langle \sum_{i=1}^{N} c_i k(., x_i), \sum_{j=1}^{N} c_j k(., x_j)\rangle_{\mathcal{H}} \\
&= \| \sum_{i=1}^{N} c_i k(., x_i)\|_{\mathcal{H}}^2 \\
&\geq 0
\end{aligned}$$

That is

Elements of the RKHS are real-valued functions on E of the form $f(.) = \sum_{i=1}^{N} c_i k(., x_i)$.

Positive Definite Kernel

Examples of reproducing kernels or positive definite kernels

- Linear kernel $k(x, y) = \langle x, y \rangle$ $x, y \in \mathbb{R}^D$

Positive Definite Kernel

Examples of reproducing kernels or positive definite kernels

- Linear kernel $k(x, y) = \langle x, y \rangle \; x, y \in \mathbb{R}^D$
- Polynomial kernel $k(x, y) = (\langle x, y \rangle + 1)^D \; x, y \in \mathbb{R}^D$

Positive Definite Kernel

Examples of reproducing kernels or positive definite kernels

- Linear kernel $k(x, y) = \langle x, y \rangle \; x, y \in \mathbb{R}^D$
- Polynomial kernel $k(x, y) = (\langle x, y \rangle + 1)^D \; x, y \in \mathbb{R}^D$
- Gaussian kernel $k(x, y) = \exp(-\|x - y\|^2/\sigma^2) \; x, y \in \mathbb{R}^D$

Positive Definite Kernel

Examples of reproducing kernels or positive definite kernels

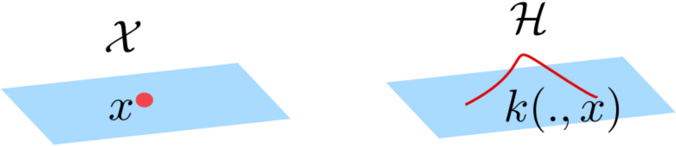
- Linear kernel $k(x, y) = \langle x, y \rangle \; x, y \in \mathbb{R}^D$
- Polynomial kernel $k(x, y) = (\langle x, y \rangle + 1)^D \; x, y \in \mathbb{R}^D$
- Gaussian kernel $k(x, y) = \exp(-\|x - y\|^2 / \sigma^2) \; x, y \in \mathbb{R}^D$
- Probability product kernel $\tilde{k}(\mathbb{P}, \mathbb{Q}) = \int_{\mathcal{X}} \mathbb{P}(x)^\rho \mathbb{Q}(x)^\rho dx$

Positive Definite Kernel

Examples of reproducing kernels or positive definite kernels

- Linear kernel $k(x, y) = \langle x, y \rangle \; x, y \in \mathbb{R}^D$
- Polynomial kernel $k(x, y) = (\langle x, y \rangle + 1)^D \; x, y \in \mathbb{R}^D$
- Gaussian kernel $k(x, y) = \exp(-\|x - y\|^2/\sigma^2) \; x, y \in \mathbb{R}^D$
- Probability product kernel $\tilde{k}(\mathbb{P}, \mathbb{Q}) = \int_{\mathcal{X}} \mathbb{P}(x)^\rho \mathbb{Q}(x)^\rho dx$
- Kernel on probability measures for $X \sim \mathbb{P}, X' \sim \mathbb{Q}$, $\tilde{k}(\mathbb{P}, \mathbb{Q}) = \langle \mathbb{E}_{\mathbb{P}}[k(X.,)], \mathbb{E}_{\mathbb{Q}}[k(X'.,)] \rangle_{\mathcal{H}}$

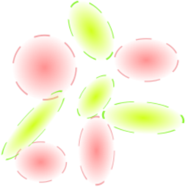
$$\boxed{\begin{array}{l}\mathcal{X} \to \mathcal{H} \\ x \mapsto k(.,x)\end{array}}$$



$$k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$$

Table of Contents

$\mathcal{X} \quad \longrightarrow \quad \mathcal{H}$



Now the problem is different

-fuzzy sets for imprecise data

-PD kernels on fuzzy sets

Uncertain and imprecise data

Kernel on fuzzy sets

Observation

Capital letters A, B, C denote sets and X, Y, Z denote fuzzy sets.

Kernel on fuzzy sets

Observation

Capital letters A, B, C denote sets and X, Y, Z denote fuzzy sets.

Remark

Notation $\mathcal{F}(\mathcal{S} \subset \Omega)$ stands for the set of all fuzzy sets over Ω whose support belongs to $\mathcal{S}$, i.e.,

$$\mathcal{F}(\mathcal{S} \subset \Omega) = \{X \subset \Omega | X_{>0} \in \mathcal{S}\}.$$

where $\mathcal{S}$ is a semi-ring of sets on Ω

Observation

Capital letters A, B, C denote sets and X, Y, Z denote fuzzy sets.

Remark

Notation $\mathcal{F}(\mathcal{S} \subset \Omega)$ stands for the set of all fuzzy sets over Ω whose support belongs to $\mathcal{S}$, i.e.,

$$\mathcal{F}(\mathcal{S} \subset \Omega) = \{X \subset \Omega | X_{>0} \in \mathcal{S}\}.$$

where $\mathcal{S}$ is a semi-ring of sets on Ω

Example

If $X \cap Y \in \mathcal{F}(\mathcal{S} \subset \Omega)$ then satisfy (finite decomposition):

$$(X \cap Y)_{>0} = \bigcup_{i \in I} A_i, \ A_i \in \mathcal{S},$$

Kernel on fuzzy sets

Example cont.

We can measure $(X \cap Y)_{>0} = \bigcup_{i\in I} A_i,\ A_i \in \mathcal{S}$ using the measure $\rho : \mathcal{S} \to [0, \infty]$ as follows:

$$\rho((X \cap Y)_{>0}) = \rho(\bigcup_{i\in I} A_i) = \sum_{i\in I} \rho(A_i),$$

Kernel on fuzzy sets

Definition (Kernels on Fuzzy Sets)

A kernel on fuzzy sets is a real valued-function

$$\begin{aligned} k : \mathcal{F}(\mathcal{S} \subset \Omega) \times \mathcal{F}(\mathcal{S} \subset \Omega) &\rightarrow \mathbb{R} \\ (\ X \quad , \quad Y\ ) &\mapsto k(X, Y), \end{aligned} \tag{9}$$

Kernel on fuzzy sets

Definition (Kernels on Fuzzy Sets)

A kernel on fuzzy sets is a real valued-function

$$\begin{aligned} k : \mathcal{F}(\mathcal{S} \subset \Omega) \times \mathcal{F}(\mathcal{S} \subset \Omega) &\rightarrow \mathbb{R} \\ (\ X \quad , \quad Y\ ) &\mapsto k(X, Y), \end{aligned} \tag{9}$$

Observation

Because each fuzzy set X belongs to $\mathcal{F}(\mathcal{S} \subset \Omega)$, then the support $X_{>0}$ of X admits a finite decomposition, that is,

$$X_{>0} = \bigcup_{i \in I} A_i, \ A_i \in \mathcal{S}$$

where $\{A_1, A_2, \dots, A_N\}$ are pairwise disjoint sets and I stand for an arbitrary index set.

Intersection Kernel on Fuzzy Sets

Definition (Intersection Kernel on Fuzzy Sets)

Let X, Y in $\mathcal{F}(\mathcal{S} \subset \Omega)$, the intersection kernel on fuzzy sets is

$$\begin{aligned} k : \mathcal{F}(\mathcal{S} \subset \Omega) \times \mathcal{F}(\mathcal{S} \subset \Omega) &\rightarrow \mathbb{R} \\ (X \quad , \quad Y) &\mapsto k(X, Y) = g(X \cap Y), \end{aligned}$$

Intersection Kernel on Fuzzy Sets

Definition (Intersection Kernel on Fuzzy Sets)

Let X, Y in $\mathcal{F}(\mathcal{S} \subset \Omega)$, the intersection kernel on fuzzy sets is

$$\begin{array}{rcl} k : \mathcal{F}(\mathcal{S} \subset \Omega) \times \mathcal{F}(\mathcal{S} \subset \Omega) & \rightarrow & \mathbb{R} \\ (\ X \quad , \quad Y\ ) & \mapsto & k(X, Y) = g(X \cap Y), \end{array}$$

Where

$$\begin{array}{rcl} g : \mathcal{F}(\mathcal{S} \subset \Omega) & \rightarrow & [0, \infty] \\ X & \mapsto & g(X) \end{array}$$

Intersection Kernel on Fuzzy Sets

Definition (Intersection Kernel on Fuzzy Sets)

Let X, Y in $\mathcal{F}(\mathcal{S} \subset \Omega)$, the intersection kernel on fuzzy sets is

$$\begin{array}{rcl} k : \mathcal{F}(\mathcal{S} \subset \Omega) \times \mathcal{F}(\mathcal{S} \subset \Omega) & \rightarrow & \mathbb{R} \\ (\; X \quad , \quad Y\;) & \mapsto & k(X, Y) = g(X \cap Y), \end{array}$$

Where

$$\begin{array}{rcl} g : \mathcal{F}(\mathcal{S} \subset \Omega) & \rightarrow & [0, \infty] \\ X & \mapsto & g(X) \end{array}$$

and the FS $X \cap Y \in \mathcal{F}(\mathcal{S} \subset \Omega)$ has M.F.

$$\mu_{X \cap Y} : \Omega \rightarrow [0, 1] \tag{10}$$

$$x \mapsto \mu_{X \cap Y} = T(\mu_X(x), \mu_Y(x)) \tag{11}$$

Kernel on fuzzy sets

From previous example.

We can measure $(X \cap Y)_{>0} = \bigcup_{i \in I} A_i, \; A_i \in \mathcal{S}$ using the measure $\rho : \mathcal{S} \to [0, \infty]$ as follows:

$$\rho((X \cap Y)_{>0}) = \rho(\bigcup_{i \in I} A_i) = \sum_{i \in I} \rho(A_i),$$

Kernel on fuzzy sets

From previous example.

We can measure $(X \cap Y)_{>0} = \bigcup_{i \in I} A_i,\ A_i \in \mathcal{S}$ using the measure $\rho : \mathcal{S} \to [0, \infty]$ as follows:

$$\rho((X \cap Y)_{>0}) = \rho(\bigcup_{i \in I} A_i) = \sum_{i \in I} \rho(A_i),$$

Adding fuzziness

The idea to include fuzziness is to weight each $\rho(A_i)$ by a value given by the contribution of the membership function on all the elements of the set A_i.

Intersection Kernel on Fuzzy Sets

Definition (Intersection Kernel on Fuzzy Sets with measure ρ)

Using $(X \cap Y)_{>0} = \bigcup_{i \in I} A_i,\ A_i \in \mathcal{S}$. Let g be the function

$$\begin{aligned} g : \mathcal{F}(\mathcal{S} \subset \Omega) &\rightarrow [0, \infty] \\ X \cap Y &\mapsto g(X \cap Y) = \sum_{i \in I} \mu_{X \cap Y}(A_i)\rho(A_i) \end{aligned}$$

where

$$\mu_{X \cap Y}(A_i) = \sum_{x \in A_i} \mu_{X \cap Y}(x).$$

Intersection Kernel on Fuzzy Sets

Definition (Intersection Kernel on Fuzzy Sets with measure ρ)

Using $(X \cap Y)_{>0} = \bigcup_{i \in I} A_i, \ A_i \in \mathcal{S}$. Let g be the function

$$
\begin{array}{rcl}
g : \mathcal{F}(\mathcal{S} \subset \Omega) & \rightarrow & [0, \infty] \\
X \cap Y & \mapsto & g(X \cap Y) = \sum_{i \in I} \mu_{X \cap Y}(A_i) \rho(A_i)
\end{array}
$$

where

$$\mu_{X \cap Y}(A_i) = \sum_{x \in A_i} \mu_{X \cap Y}(x).$$

Intersection Kernel on Fuzzy Sets with measure ρ

We define the *Intersection Kernel on Fuzzy Sets with measure* ρ as:

$$
\begin{aligned}
k(X, Y) &= g(X \cap Y) \\
&= \sum \mu_{X \cap Y}(A_i) \rho(A_i) \qquad (12)
\end{aligned}
$$

Using the T-norm operator, the intersection kernel on fuzzy sets with measure ρ can be written as:

$$\begin{aligned} k(X,Y) &= \sum_{i \in I} \mu_{X \cap Y}(A_i)\rho(A_i) \\ &= \sum_{i \in I} \sum_{x \in A_i} \mu_{X \cap Y}(x)\rho(A_i) \\ &= \sum_{i \in I} \sum_{x \in A_i} T(\mu_X(x), \mu_Y(x))\rho(A_i) \end{aligned}$$

Some kernel examples for different T-norm operators

Intersection kernels on fuzzy sets with measure ρ	
$k_{\min}(X,Y)$	$\sum_{i\in I}\sum_{x\in A_i}\min(\mu_X(x),\mu_Y(x))\rho(A_i)$
$k_P(X,Y)$	$\sum_{i\in I}\sum_{x\in A_i}\mu_X(x)\mu_Y(x)\rho(A_i)$
$k_{\max}(X,Y)$	$\sum_{i\in I}\sum_{x\in A_i}\max(\mu_X(x)+\mu_Y(x)-1,0)\rho(A_i)$
$k_Z(X,Y)$	$\sum_{i\in I}\sum_{x\in A_i}Z(\mu_X(x),\mu_Y(x))\rho(A_i)$

Table: kernels on fuzzy sets.

Function Z is defined as

$$Z(\mu_X(x),\mu_Y(x)) = \begin{cases} \mu_X(x), \text{ if } \mu_Y(x)=1 \\ \mu_Y(x), \text{ if } \mu_X(x)=1 \\ 0, \textit{ otherwise} \end{cases}$$

Lema

$k_{\min}(X, Y) = \sum_{i \in I} \sum_{x \in A_i} \min(\mu_X(x), \mu_Y(x))\rho(A_i)$
is positive definite

Lema

$k_P(X, Y) = \sum_{i \in I} \sum_{x \in A_i} \mu_X(x)\mu_Y(x)\rho(A_i)$
is positive definite.

It is worth to note that, if the σ-algebra is a Borel algebra of subsets of $\mathbb{R}^D$, then the intersection kernel with measure ρ can be written as

$$k(X, Y) = \int_{\mathbb{R}^D} T(\mu_X(x), \mu_Y(x))d\rho(x)$$

as for example k_{min} and k_P can be written as

$$k_{min}(X, Y) = \int_{\mathbb{R}^D} \min(\mu_X(x), \mu_Y(x))d\rho(x) \quad (13)$$

$$k_P = \int_{\mathbb{R}^D} \mu_X(x)\mu_Y(x)d\rho(x) \quad (14)$$

$$(15)$$

Definition

Let $k : \Omega \times \Omega \to \mathbb{R}$ be a positive definite kernel. The cross product kernel between fuzzy sets $X, Y \in \mathcal{F}(\mathcal{S} \subset \Omega)$ is the real valued function $k_\times$ defined on $\mathcal{F}(\mathcal{S} \subset \Omega) \times \mathcal{F}(\mathcal{S} \subset \Omega)$ as

$$k_\times(X, Y) = \sum_{x \in X} \sum_{y \in Y} k(x, y) \mu_X(x) \mu_Y(y) \tag{16}$$

Lema

kernel $k_\times$ is positive definite

Definition (Nonsingleton TSK Fuzzy Kernel)

Let $X \cap Y$ be a fuzzy set given by Definition (3.2) and let g be the function:

$$\begin{aligned} g : \mathcal{F}(\mathcal{S} \subset \Omega) &\rightarrow [0, \infty] \\ X \cap Y &\mapsto g(X \cap Y) = \sup_{x \in \Omega} \mu_{X \cap Y}(x), \end{aligned}$$

then the *Nonsingleton TSK Fuzzy Kernel* is given by :

$$k_{tks}(X, Y) = \sup_{x \in \Omega} \mu_{X \cap Y}(x) \tag{17}$$

Using T-norm operators, this kernel can be written as:

$$\begin{aligned} k_{tks}(X, Y) &= \sup_{x \in \Omega} \mu_{X \cap Y}(x) \\ &= \sup_{x \in \Omega} T(\mu_X(x), \mu_Y(x)) \end{aligned}$$

Lema

The Nonsingleton TSK Fuzzy Kernel is positive definite, that is:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k_{tks}(X_i, X_j) \geq 0,$$

$\forall N \in \mathbb{N},\ \forall c_i, c_j \in \mathbb{R},\ \forall X_i, X_j \in \mathcal{F}(\mathcal{S} \subset \Omega).$

Examples of nonsingleton TSK fuzzy kernels

Gaussian MF

$$
\begin{aligned}
k_{tks}(X, Y) &= \sup_{x \in \Omega} \mu_{X \cap Y}(x) \\
&= \sup_{x \in \Omega} T(\mu_X(x), \mu_Y(x)) \\
&= \exp\left\{ -\frac{1}{2} \sum_{j=1}^{p} \frac{(m_j - m_j^l)^2}{\sigma_j^2 + (\sigma_j^l)^2} \right\}
\end{aligned}
$$

Examples of nonsingleton TSK fuzzy kernels

Gaussian MF

$$\begin{aligned} k_{tks}(X, Y) &= \sup_{x \in \Omega} \mu_{X \cap Y}(x) \\ &= \sup_{x \in \Omega} T(\mu_X(x), \mu_Y(x)) \\ &= \exp\left\{ -\frac{1}{2} \sum_{j=1}^{p} \frac{(m_j - m_j^l)^2}{\sigma_j^2 + (\sigma_j^l)^2} \right\} \end{aligned}$$

Gaussian MF with parameter $\gamma \in \mathbb{R}$

$$k_{tks,\gamma}(X, Y) = \exp\left\{ -\frac{1}{2} \sum_{j=1}^{p} \frac{(m_j - m_j^l)^2}{\sigma_j^2 + (\sigma_j^l)^2 + \gamma} \right\} \quad (18)$$

More kernels on fuzzy sets

If $k_1(.,.)$ and $k_2(.,.)$ are PD kernels on fuzzy sets, by closure properties of PD kernels, also are PD kernels on fuzzy sets:

1. $k_1(X,Y) + k_2(X,Y)$;
2. $\alpha k_1(X,Y), \quad \alpha \in \mathbb{R}^+$;
3. $k_1(X,Y)k_2(X,Y)$;
4. $f(X)f(Y), \quad f : \mathcal{F}(\mathcal{S} \subset \Omega) \to \mathbb{R}$;
5. $k_1(f(X),f(Y)), \quad f : \mathcal{F}(\mathcal{S} \subset \Omega) \to \mathcal{F}(\mathcal{S} \subset \Omega)$;
6. $\exp(k_1(X,Y))$;
7. $p(k_1(X,Y)), \quad p$ is a polynomial with positive coefficients.

More kernels on fuzzy sets

More kernels on fuzzy sets could be obtained using the nonlinear mapping

$$\begin{aligned} \phi : \mathcal{F}(\mathcal{S} \subset \Omega) &\rightarrow \mathcal{H} \\ X &\mapsto \phi(X), \end{aligned}$$

and using the fact that $k(X, Y) = \langle \phi(X), \phi(Y) \rangle_{\mathcal{H}}$ and

$$\begin{aligned} D(X, Y) &\stackrel{\text{def}}{=} \|\phi(X) - \phi(Y)\|^2_{\mathcal{H}} \\ &= k(X, Y) - 2k(X, Y) + k(Y, Y), \end{aligned}$$

The following kernels are PD kernels on fuzzy sets:

- **Fuzzy Polynomial kernel** $\alpha \geq 0, \beta \in \mathbb{N}$

$$\begin{aligned} k_{pol}(X, Y) &= (\langle \phi(X), \phi(Y) \rangle_{\mathcal{H}} + \alpha)^{\beta} \\ &= (k(X, Y) + \alpha)^{\beta}. \end{aligned}$$

- **Fuzzy Gaussian kernel** $\gamma > 0$

$$\begin{aligned} k_{gauss}(X, Y) &= \exp(-\gamma \|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2) \\ &= \exp(-\gamma D(X, Y)). \end{aligned}$$

- **Fuzzy Rational Quadratic kernel** $\alpha, \beta > 0$

$$\begin{aligned} k_{ratio}(X, Y) &= (1 + \frac{\|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2}{\alpha\beta^2})^{-\alpha} \\ &= (1 + \frac{D(X, Y)}{\alpha\beta^2})^{-\alpha}. \end{aligned}$$

Conditionally Positive Definite Kernels on Fuzzy Sets

Lema

CPD kernels are symmetric kernels satisfying:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(x_i, x_j) \geq 0, \; \sum_{i=1}^{N} c_i = 0. \tag{19}$$

$\forall N \in \mathbb{N}, \; \forall c_i, c_j \in \mathbb{R}.$

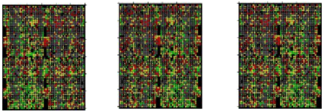
- **Fuzzy Multiquadric kernel**

$$\begin{aligned} k_{multi}(X, Y) &= -\sqrt{\|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2 + \alpha^2} \\ &= -\sqrt{D(X, Y) + \alpha^2}. \end{aligned}$$

- **Fuzzy Inverse Multiquadric kernel**

$$k_{invmult}(X, Y) = (\sqrt{\|\phi(X) - \phi(Y)\|_{\mathcal{H}}^2 + \alpha^2})^{-1}$$

Table of Contents

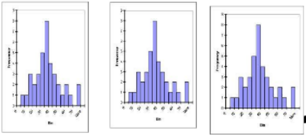


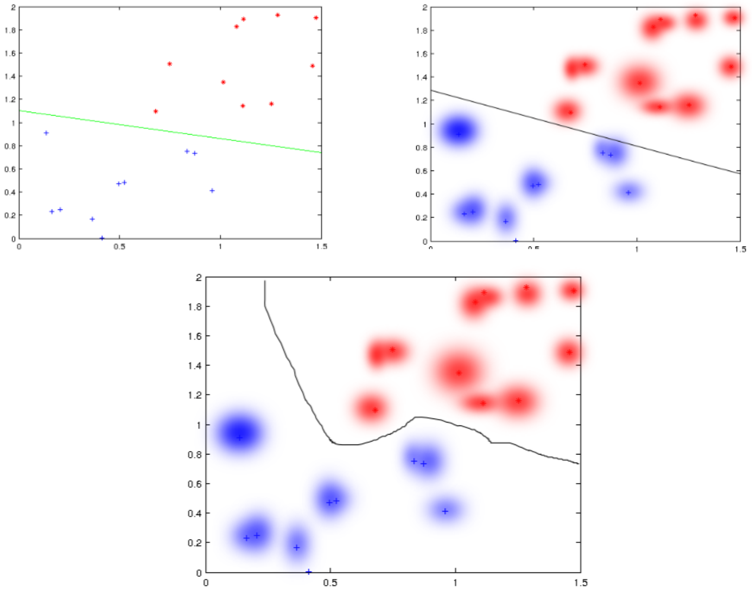
gene expression data

subjective data

interval data

histogram data




2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5

Machine Learning Applications

- ML Applications in fuzzy data, i.e., $\{X_i\}_{i=1}^N$, where each X_i is FS.

Machine Learning Applications

- ML Applications in fuzzy data, i.e., $\{X_i\}_{i=1}^{N}$, where each X_i is FS.
- Kernels on FS can be used modularly in SVM, SVDD, kernel PCA, Gaussian process.

Machine Learning Applications

- ML Applications in fuzzy data, i.e., $\{X_i\}_{i=1}^{N}$, where each X_i is FS.
- Kernels on FS can be used modularly in SVM, SVDD, kernel PCA, Gaussian process.
- ML on big data, by data squashing.

Machine Learning Applications

- ML Applications in fuzzy data, i.e., $\{X_i\}_{i=1}^N$, where each X_i is FS.
- Kernels on FS can be used modularly in SVM, SVDD, kernel PCA, Gaussian process.
- ML on big data, by data squashing.
- Promising results on low quality data.

Reference

J. Guevara, R. Hirata, and S. Canu, "Kernel functions in Takagi-Sugeno-Kang fuzzy system with nonsingleton fuzzy input," in Fuzzy Systems (FUZZ), 2013 IEEE International Conference on, 2013, pp. 1-8.

Table of Contents

Conclusion

- Theory of kernels allow us to represent fuzzy sets as functions in a RKHS.
- Tool to use: semiring of sets and measures
- Important applications on fuzzy data

Summary

Summary

- kernels on fuzzy sets

Summary

- kernels on fuzzy sets
- PD fuzzy intersection kernels

Summary

- kernels on fuzzy sets
- PD fuzzy intersection kernels
- CPD fuzzy kernels

Acknowledge

The authors are thankful with FAPESP grant # 2011/50761-2, CNPq, CAPES, NAP eScience - PRP - USP for their financial support.