Institut Charles Delaunay

Regularization and reproducing kernels

Pattern Recognition
OS14

- Chapitre 2 -

Learning problem :

We are looking for a function ψ of a function space $\mathcal{H}$ which contains candidate functions from $\mathcal{X}$ to $\mathcal{Y}$. For each $\boldsymbol{x} \in \mathcal{X}$, these functions predict a corresponding label y. Thus :

$$y = \psi(\boldsymbol{x})$$

We have a training set $\mathcal{A}_n = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$

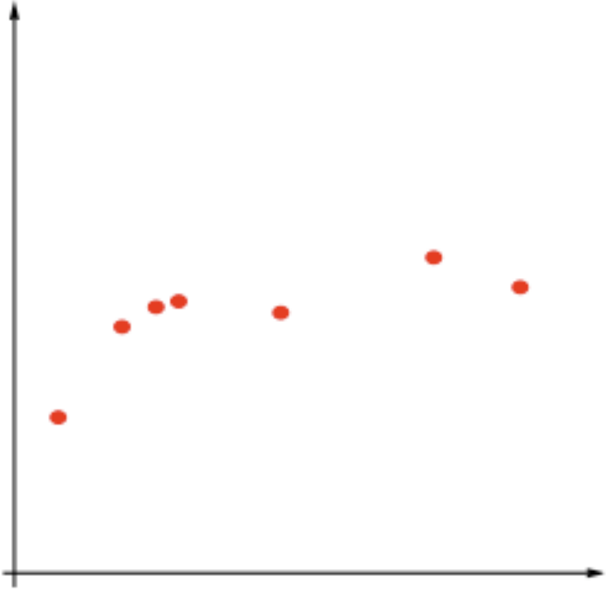
$\triangleright$ *Empirical risk minimization and generalization !*

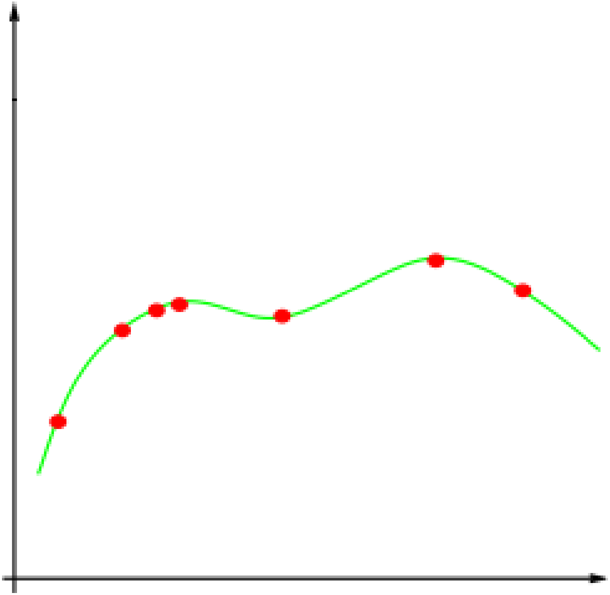
Definition (Well-posed Problem / Ill-posed Problem (Hadamard))

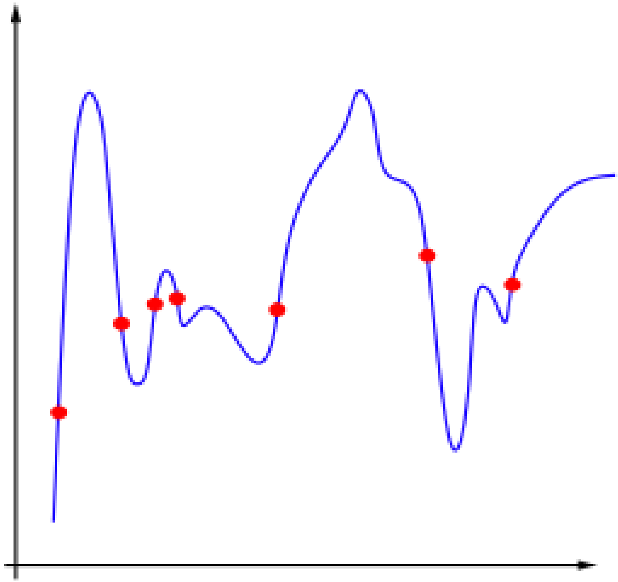
A problem is *well-posed* if

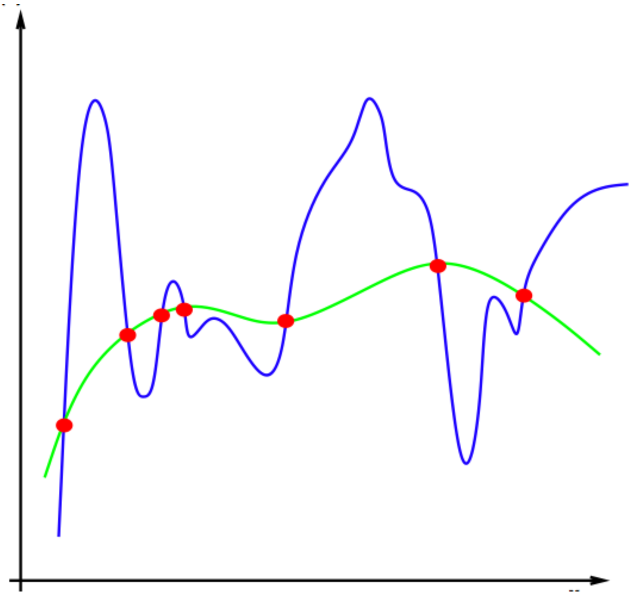
- it has a solution
- the solution is unique
- the solution is a continuous function of the data. (a small perturbation of the data produces a small perturbation of the solution)

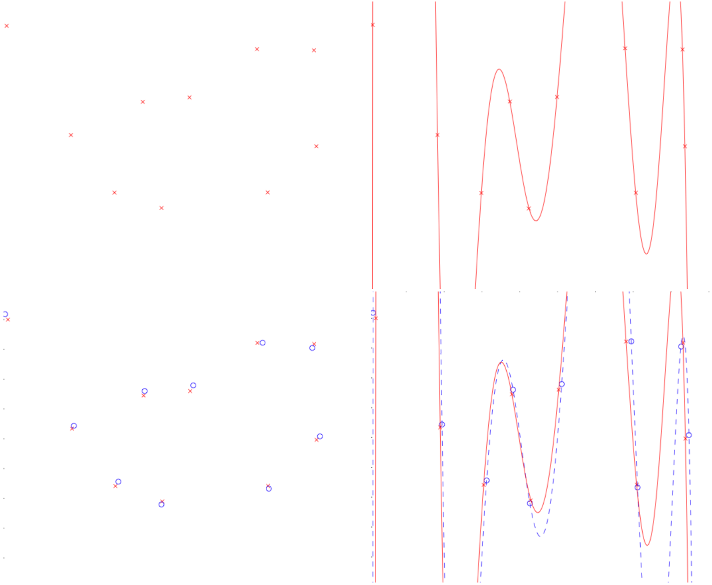
A problem is *ill-posed* if it is not well-posed...

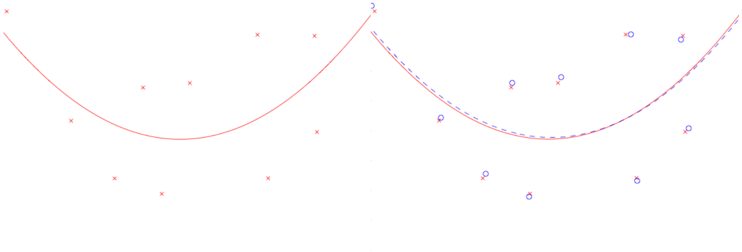












The empirical risk minimization

$$J_{emp}(\psi) = \frac{1}{n}\sum_{k=1}^{n} Q(\psi(\boldsymbol{x}_k), y_k),$$

is an ill-posed Problem.

Solution : **Regularization**

Ivanov regularization

Find the function ψ which minimizes

$$\frac{1}{n}\sum_{k=1}^{n} Q(\psi(\boldsymbol{x}_k), y_k),$$

under the constraint :

$$\|\psi\|^2 \leq A$$

Empirical risk penalization :

$$\mathsf{RisqEmp}(\psi) + \eta \ \mathsf{Pen}(\psi),$$

with $\eta \in \mathbb{R}^{+*}$ which is a positive parameter controlling the tradeoff between these two terms.

▷ *The penalty term is used to incorporate a smoothing effect*

Tikhonov regularization

Determine the function ψ in a space $\mathcal{H}$ of candidate functions, which minimizes :

$$\frac{1}{n}\sum_{k=1}^{n} Q(\psi(\boldsymbol{x}_k), y_k) + \eta\|\psi\|^2_{\mathcal{H}},$$

for a parameter $\eta > 0$, and where $\|\psi\|_{\mathcal{H}}$ is the function norm in the space $\mathcal{H}$.

This problem is well-posed.

Problem

Given $\mathcal{A}_n = \{(X_i, Y_i)\}_{i=1}^{n}$ with $X_i \in \mathbb{R}^d$ and $Y \in \mathbb{R}$ a training set.

Find a linear regression $\hat{y} = \boldsymbol{a}^T \boldsymbol{x}$ with :

$$Q(\psi(\boldsymbol{x}_k), y_k) = \left(y_k - \boldsymbol{a}^T \boldsymbol{x}_k - a_0\right)^2$$

and

$$\|\psi\|_{\mathcal{H}}^2 = \|[\boldsymbol{a}\ a_0]\|^2$$

Define :

- $\tilde{\boldsymbol{x}} = [\boldsymbol{x}\ 1]^T$
- $\tilde{\boldsymbol{a}} = [\boldsymbol{a}\ a_0]^T$
- $\mathrm{X} = [\tilde{\boldsymbol{x}}_1 \tilde{\boldsymbol{x}}_2 \ldots \tilde{\boldsymbol{x}}_n] \in \mathbb{R}^{(d+1)\times n}$
- $\mathrm{y} = [y_1 y_2 \ldots y_n]^T \in \mathbb{R}^n$

Define $\eta > 0$.

Formulation

Find :

$$\begin{aligned} \tilde{\boldsymbol{a}}^* &= \underset{\tilde{\boldsymbol{a}} \in R^{d+1}}{\arg\min} \left(\sum_{i=1}^{n} \left(y_i - \tilde{\boldsymbol{a}}^T \tilde{x}_i\right)^2 + \eta \left\|\tilde{\boldsymbol{a}}\right\|^2 \right) \\ &= \underset{\tilde{\boldsymbol{a}} \in R^{d+1}}{\arg\min} \left(\left\|\mathrm{y}^T - \tilde{\boldsymbol{a}}^T \tilde{\mathrm{X}}\right\|^2 + \eta \left\|\tilde{\boldsymbol{a}}\right\|^2 \right) \end{aligned}$$

Solution

$$\tilde{a}^* = \left(\tilde{\mathrm{X}}\tilde{\mathrm{X}}^T + \eta \mathrm{I}\right)^{-1} \tilde{\mathrm{X}}\mathrm{y}$$

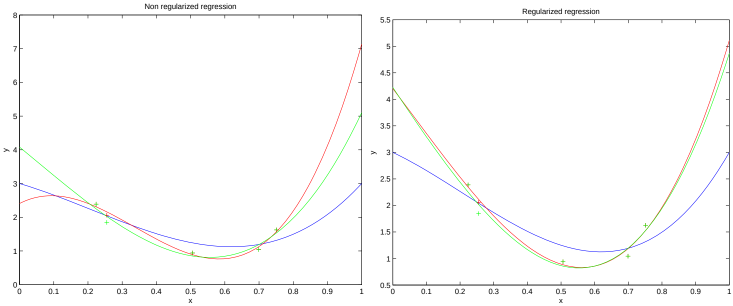


Figure: True function (blue), regressions using cubic polynomial with 2 datasets (red and green) - no regularization (letf) - regularization (right)

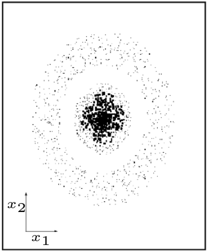
Reproducing Kernel Hilbert Space

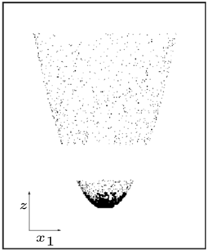
Intuitions

1. Simplest function ψ : linear
2. Role of η : How far from $\psi_i = C^{te}$

Space $\mathcal{H}$

Map to higher dimension : Feature space
Example : $\phi(x) = [x_1 \ x_2 \ x_1^2 + x_2^2] \in \mathbb{R}$





Outline

1. Hilbert space
2. Kernels
3. Reproducing property

Definition

$\mathcal{H}$ is an Hilbert space if :

1. $\mathcal{H}$ is a functional space
2. with a dot product
3. and which has a norm deduced from the dot product that enable to measure distances.

Definition - dot product

Let $\mathcal{H}$ be a functional space over $\mathbb{R}$. A function $\langle .,. \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is a dot product on $\mathcal{H}$ if :

- Linear : $\langle \alpha\psi, \phi \rangle_{\mathcal{H}} = \alpha \langle \phi, \psi \rangle_{\mathcal{H}}$ and $\langle \psi_1 + \psi_2, \phi \rangle_{\mathcal{H}} = \langle \psi_1, \phi \rangle_{\mathcal{H}} + \langle \psi_2, \phi \rangle_{\mathcal{H}}$
- Symmetric : $\langle \psi, \phi \rangle_{\mathcal{H}} = \langle \phi, \psi \rangle_{\mathcal{H}}$
- Positive : $\langle f, f \rangle_{\mathcal{H}} \geq 0$ and $\langle f, f \rangle_{\mathcal{H}} = 0 \quad \Rightarrow \quad f = 0$

Norm

$\|f\|_{\mathcal{H}} = \sqrt{\langle f, f \rangle_{\mathcal{H}}}$

It is a generalization of Euclidian Space.
The dimension of an Hilbert space can be infinite.

Properties of norm

The *norm* $\|\cdot\|_{\mathcal{H}}$ defined on a space $\mathcal{H}$ is an application from $\mathcal{H}$ in $\mathbb{R}$, nonnegative, which satisfy the following properties, for any $\psi, \phi \in \mathcal{H}$,

- Positivity : $\|\psi\|_{\mathcal{H}} \geq 0$, with equality if and only if $\psi = 0$
- Homogeneity : $\|\alpha\psi\|_{\mathcal{H}} = |\alpha| \; \|\psi\|_{\mathcal{H}}$ for any $\alpha \in \mathbb{R}$
- Triangular inequality : $\|\psi + \phi\|_{\mathcal{H}} \leq \|\psi\|_{\mathcal{H}} + \|\phi\|_{\mathcal{H}}$

Example : $\mathcal{L}_2[a,b] = \{\psi \mid \int_a^b |\psi^2(x)|dx < \infty\}$ is a Hilbert space where the dot product is defined by :

$$\langle \psi, \phi \rangle = \int_a^b \psi(x)\phi(x)dx$$

A Hilbert space is a (possibly) infinite dimensional vector space endowed with a dot product.

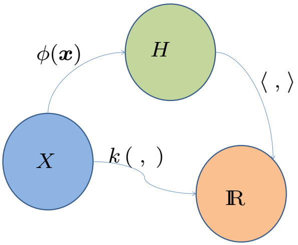
Definition

Let $\mathcal{X}$ be a representation space.
A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ **is a kernel** if there exists :

- $\mathcal{H}$, an $\mathbb{R}$-Hilbert space and
- a mapping function $\phi : \mathcal{X} \to \mathcal{H}$

such that :

$$\forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \quad \kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \left\langle \phi\left(\boldsymbol{x}\right), \phi\left(\boldsymbol{x}'\right) \right\rangle_{\mathcal{H}}$$



- Almost no conditions on $\mathcal{X}$
  No need for dot product
  Documents
  Structured data
- A kernel can correspond to several mapping functions :
  Example : $\mathcal{X}$ is $\mathbb{R}$:

$$\phi_1(x) = x$$

$$\phi_2(x) = \begin{bmatrix} \frac{1}{\sqrt{2}}x \\ \frac{1}{\sqrt{2}}x \end{bmatrix}$$

Sum

Let κ_1 and κ_2 be two kernels on $\mathcal{X}$, then $\kappa_1 + \kappa_2$ is also a kernel on $\mathcal{X}$.

A difference of kernel may not be a kernel.

Product

Let $\alpha > 0$ a real and κ a kernel on $\mathcal{X}$, then $\alpha\kappa$ is also a kernel in $\mathcal{X}$.

Mapping

Consider :

- two representation spaces $\mathcal{X}_1$ and $\mathcal{X}_2$
- a mapping $\eta : \mathcal{X}_1 \to \mathcal{X}_2$
- a kernel $\kappa_2 : \mathcal{X}_2 \times \mathcal{X}_2 \to \mathbb{R}$
- $\boldsymbol{x}$ and $\boldsymbol{x}'$ elements of $\mathcal{X}_1$

then $\kappa_2\left(\eta(\boldsymbol{x}), \eta(\boldsymbol{x}')\right)$ is a kernel on $\mathcal{X}_1$

Example

Product of kernels

Consider :

- two representation spaces $\mathcal{X}_1$ and $\mathcal{X}_2$
- two kernels κ_1 and κ_2

then $\kappa = \kappa_1 \times \kappa_2$ is a kernel on $\mathcal{X}_1 \times \mathcal{X}_2$

If $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$ then κ is also kernel on $\mathcal{X}$.
Interest

Consequence

Consider $\boldsymbol{x}$ and $\boldsymbol{x}' \in \mathbb{R}^d$, $m \in \mathbb{N}^*$ and $a \in \mathbb{R}^+$:

$$\kappa\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \left(\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + a\right)^m$$

is a valid kernel.

Can a kernel be a dot product of an infinity of features ?

The $\mathcal{L}_2$ norm of the kernel must be bounded.

Let $\phi_i(\boldsymbol{x})$ be the i^{th} coordinate of $\phi(\boldsymbol{x})$ then

$$\begin{aligned}\kappa\left(\boldsymbol{x},\boldsymbol{x}'\right) &= \left\langle \phi\left(\boldsymbol{x}\right), \phi\left(\boldsymbol{x}'\right)\right\rangle \\ &= \sum_{i=1}^{\infty} \phi_i\left(\boldsymbol{x}\right)\phi_i\left(\boldsymbol{x}'\right) \\ &\leq \left\|\phi\left(\boldsymbol{x}\right)\right\| \left\|\phi\left(\boldsymbol{x}'\right)\right\|\end{aligned}$$

which is bounded if the sequence of $\left(\phi_i(\boldsymbol{x})\right)_{i\geq 0}$ are in the space ℓ_2.
Space ℓ_2 is the set of all sequences squared summable.

Example

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = exp\left(\langle \boldsymbol{x}, \boldsymbol{x}' \rangle\right)$$

Proof

Du to Cauchy-Schwarz :

$$\left|\langle \boldsymbol{x}, \boldsymbol{x}' \rangle\right| \leq \|\boldsymbol{x}\| \left\|\boldsymbol{x}'\right\|$$

Which is bounded
and

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=0}^{\infty} a_i \left\langle \boldsymbol{x}, \boldsymbol{x}' \right\rangle^i$$

How can we find good kernel among all functions of 2 arguments ?

1. Try to find the mapping $\phi()$?
   - No obvious
   - May be infinite dimensional
   - The feature mapping is not unique...
2. Prove that the kernel is definite positive !

Definition :

A symmetric function is said *definite positive* if

$$\sum_{i,j} \alpha_i \alpha_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0$$

for any $\alpha_i, \alpha_j \in \mathbb{R}$ and $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{X}$.

A function $\kappa(.,.)$ is *strictly definite positive* if, for distinct $\boldsymbol{x}_i$, the equality holds only if all α_i equal 0.

The dot product of any mapping function is positive definite

Proof

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\kappa\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left\langle\alpha_i\phi\left(\boldsymbol{x}_i\right),\alpha_j\phi\left(\boldsymbol{x}_j\right)\right\rangle = \left\|\sum_{i=1}^{n}\alpha_i\phi\left(\boldsymbol{x}_i\right)\right\|^2 \geq 0$$

Important

Reverse holds !
Positive definite $\kappa(.,.)$ is a dot product in $\mathcal{H}$.

Not definite positive - and so what ?

$\Rightarrow$ Negative norm
If $\boldsymbol{v}_s$ is a eigenvector in $\mathcal{H}$ and let $\boldsymbol{z}=\sum_{i=1}^{n} v_{si}\phi(\boldsymbol{x}_i)$ then :

$$\begin{aligned}\|\boldsymbol{z}\|^2 &= \langle\boldsymbol{z},\boldsymbol{z}\rangle\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} v_{si}v_{sj}\left\langle\phi\left(\boldsymbol{x}_i\right),\phi\left(\boldsymbol{x}_j\right)\right\rangle\\ &= \boldsymbol{v}_s^T K\boldsymbol{v}_s\\ &= \lambda_s\left\|\boldsymbol{v}_s\right\|^2\end{aligned}$$

Thus all λ_s have to be positive. . .

We consider the functions $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ that can act as a dot product in a space $\mathcal{H}$. We call *kernel* a symmetric function κ of $\mathcal{X} \times \mathcal{X}$ in $\mathbb{R}$

Theorem (Mercer)

If κ is a continuous positive defined kernel based on an integral operator, which means that :

$$\iint \varphi(\boldsymbol{x})\, \kappa(\boldsymbol{x}, \boldsymbol{x}')\, \varphi^*(\boldsymbol{x}')\, d\boldsymbol{x}\, d\boldsymbol{x}' \geq 0$$

For any $\varphi \in \mathcal{L}_2(\mathcal{X})$, it can be decomposed as :

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=1}^{\infty} \lambda_i\, \psi_i(\boldsymbol{x})\, \psi_i(\boldsymbol{x}') = \langle \boldsymbol{\phi}(\boldsymbol{x}), \boldsymbol{\phi}(\boldsymbol{x}') \rangle,$$

where ψ_i and λ_i are the eigenfunctions (orthogonales) and eigenvalues (positives) of the kernel κ, respectively, such that :

$$\int \kappa(\boldsymbol{x}, \boldsymbol{x}')\, \psi_i(\boldsymbol{x})\, d\boldsymbol{x} = \lambda_i\, \psi_i(\boldsymbol{x}').$$

It is easy to see that a kernel κ satisfying Mercer's theorem can act as a scalar product in a transformed space $\mathcal{H}$.
Since :

$$\boldsymbol{\phi}(\boldsymbol{x}) = \begin{pmatrix} \sqrt{\lambda_1}\,\psi_1(\boldsymbol{x}) \\ \sqrt{\lambda_2}\,\psi_2(\boldsymbol{x}) \\ \cdots \end{pmatrix}$$

Under these conditions, it is verified that :

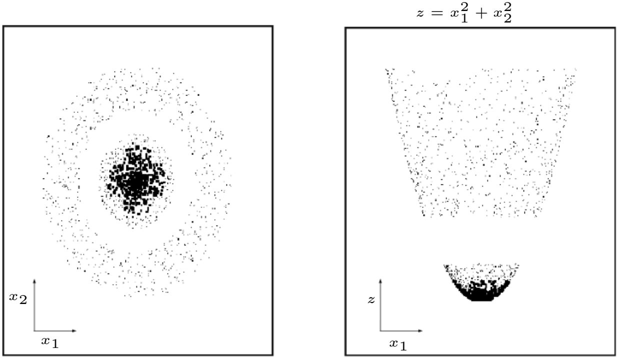
$$\langle \boldsymbol{\phi}(\boldsymbol{x}), \boldsymbol{\phi}(\boldsymbol{x}') \rangle = \kappa(\boldsymbol{x}, \boldsymbol{x}')$$

So, let define the space $\mathcal{H}$ as the space generated by the eigenfunctions ψ_i of kernel κ which means that :

$$\mathcal{H} = \{f(\cdot) \mid f(x) = \sum_{i=1}^{\infty} \alpha_i \; \psi_i(x), \; \alpha_i \in \mathbb{R}\}.$$

Property

$\phi(\boldsymbol{x})$ is often of large dimension, sometimes infinite.



The polynomial transformation makes the data linearly separable.

A linear classifier in the space defined by $\phi(\boldsymbol{x})$ is non-linear with respect to $\boldsymbol{x}$

Property

We **never** need to explicitly calculate $\boldsymbol{\phi}(\boldsymbol{x})$

In the case of the polynomial transformation of order 2, it is easy to show that :

$$\langle \boldsymbol{\phi}(\boldsymbol{x}), \boldsymbol{\phi}(\boldsymbol{x}') \rangle = (1 + \langle \boldsymbol{x}, \boldsymbol{x}' \rangle)^2 \triangleq \kappa(\boldsymbol{x}, \boldsymbol{x}')$$

$\boldsymbol{\phi}(\boldsymbol{x}) = [1, \quad \boldsymbol{x}, \quad \boldsymbol{x}, \quad \boldsymbol{x}^2]$

▷ **The dot product computation can be performed in $\mathbb{R}^2$!**

In a more general case (polynomial of order q), it generalizes to :
$\kappa(\boldsymbol{x}, \boldsymbol{x'}) = (1 + \langle \boldsymbol{\phi}(\boldsymbol{x}), \boldsymbol{\phi}(\boldsymbol{x}') \rangle)^q$, with $\boldsymbol{x} \in \mathbb{R}^l$.

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = (1 + \langle \boldsymbol{x}, \boldsymbol{x}' \rangle)^q = \sum_{j=0}^{q} \binom{q}{j} \langle \boldsymbol{x}, \boldsymbol{x}' \rangle^j .$$

Each componant $\langle \boldsymbol{x}, \boldsymbol{x}' \rangle^j = [x(1)\, x'(1) + \ldots + x(l)\, x'(l)]^j$ of this expression can be develop as a weighted sum of order j monomials :

$$[x(1)\, x'(1)]^{j_1} [x(2)\, x'(2)]^{j_2} \ \ldots \ [x(l)\, x'(l)]^{j_l}$$

with $\sum_{i=1}^{l} j_i = j$. The expression of $\boldsymbol{\phi}(\boldsymbol{x})$ can be deduced...

It can be shown that the following kernels verify the condition of Mercer, and thus correspond to a dot product in a space $\mathcal{H}$.

Projective kernels	
monomial of degree q	$\langle \boldsymbol{x}, \boldsymbol{x}' \rangle^q$
polynomial of degree q	$(1 + \langle \boldsymbol{x}, \boldsymbol{x}' \rangle)^q$
sigmoidal	$\frac{1}{\eta_0} \tanh(\beta_0 \langle \boldsymbol{x}, \boldsymbol{x}' \rangle - \alpha_0)$

Radial kernels	
Gaussien	$\exp(-\frac{1}{2\sigma_0^2} \\|\boldsymbol{x} - \boldsymbol{x}'\\|^2)$
exponential	$\exp(-\frac{1}{2\sigma_0^2} \\|\boldsymbol{x} - \boldsymbol{x}'\\|)$
uniform	$\frac{1}{\eta_0} \mathbb{1}_{\\|\boldsymbol{x}-\boldsymbol{x}'\\| \leq \beta_0}$
Epanechnikov	$\frac{1}{\eta_0} (\beta_0^2 - \\|\boldsymbol{x} - \boldsymbol{x}'\\|^2) \mathbb{1}_{\\|\boldsymbol{x}-\boldsymbol{x}'\\| \leq \beta_0}$
Cauchy	$\frac{1}{\eta_0} \frac{1}{1+\\|\boldsymbol{x}-\boldsymbol{x}'\\|^2/\beta_0^2}$

. . . and also : $\kappa_1(\boldsymbol{x}, \boldsymbol{x}') + \kappa_2(\boldsymbol{x}, \boldsymbol{x}')$, $\kappa_1(\boldsymbol{x}, \boldsymbol{x}') \cdot \kappa_2(\boldsymbol{x}, \boldsymbol{x}')$,. . .

Let define ϕ such that :

$$\begin{aligned} \phi \quad & : \mathbb{R}^2 \to \mathbb{R}^3 \\ \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix} & \mapsto \phi(\boldsymbol{x}) = \begin{bmatrix} x_1 \\ x_2 \\ x_1 x_2 \end{bmatrix} \end{aligned}$$

with kernel

$$\kappa(x, y) = \begin{bmatrix} x_1 \\ x_2 \\ x_1 x_2 \end{bmatrix}^T \begin{bmatrix} y_1 \\ y_2 \\ y_1 y_2 \end{bmatrix}$$

Let the feature space be $\mathcal{H}$.

Define a linear function f of $\boldsymbol{x}$ and $x_1 x_2$:

$$f(\boldsymbol{x}) = f_1 x_1 + f_2 x_2 + f_3 x_1 x_2$$

f is a function that maps data from $\mathbb{R}^2$ to $\mathbb{R}$.
A representation of f can be :

$$f(.) = [f_1 \ f_2 \ f_3]^T$$

$f(.)$ is the function as an object (a vector in $\mathbb{R}^3$ in that case) $f(\boldsymbol{x}) \in \mathbb{R}$ is the function value at point $\boldsymbol{x}$

$$f(\boldsymbol{x}) = f(.)^T \phi(\boldsymbol{x}) = \langle f(.), \phi(\boldsymbol{x}) \rangle_{\mathcal{H}}$$

Evaluation of f is a dot product in $\mathcal{H}$

$\phi(\boldsymbol{y})$ is a mapping from $\mathbb{R}^2$ to $\mathbb{R}^3$ that corresponds also to a function mapping from $\mathbb{R}^2$ to $\mathbb{R}$.

$$\phi(\boldsymbol{y}) = [y_1 \ y_2 \ y_1 y_2]^T = \kappa(., \boldsymbol{y})$$

Given any $\boldsymbol{y}$, there is a vector $\kappa(., \boldsymbol{y})$ in $\mathcal{H}$ such that :

$$\langle \kappa(., \boldsymbol{y}), \phi(\boldsymbol{x}) \rangle_{\mathcal{H}} = y_1 x_1 + y_2 x_2 + y_1 y_2 x_1 x_2$$

Due to symmetry :

$$\langle \kappa(., \boldsymbol{x}), \phi(\boldsymbol{y}) \rangle_{\mathcal{H}} = \langle \kappa(., \boldsymbol{y}), \phi(\boldsymbol{x}) \rangle_{\mathcal{H}}$$

One can write $\phi(\boldsymbol{x}) = \kappa(., \boldsymbol{x})$ and $\phi(\boldsymbol{y}) = \kappa(., \boldsymbol{y})$ thus

$$\langle \kappa(., \boldsymbol{x}), \phi(\boldsymbol{y}) \rangle_{\mathcal{H}} = \langle \phi(\boldsymbol{y}), \phi(\boldsymbol{x}) \rangle_{\mathcal{H}} = \kappa(\boldsymbol{x}, \boldsymbol{y})$$

This illustrates the definition of a RKHS :

The reproducing property

$$\forall \boldsymbol{x} \in \mathcal{X}, \langle f(.), \kappa(., \boldsymbol{x}) \rangle_{\mathcal{H}} = f(\boldsymbol{x})$$

or

$$\forall \boldsymbol{x} \in \mathcal{X}, \langle f(.), \phi(\boldsymbol{x}) \rangle_{\mathcal{H}} = f(\boldsymbol{x})$$

In particular

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}, \kappa(\boldsymbol{x}, \boldsymbol{y}) = \langle \kappa(., \boldsymbol{x}), \kappa(., \boldsymbol{y}) \rangle_{\mathcal{H}}$$

Note that $\mathcal{H}$ can be larger than $\phi(\boldsymbol{x})$

Example

$$[1, 1, -1] \in \mathcal{H}$$

Theorem (The representer theorem)

Any function ψ in a reproducing kernel Hilbert space $\mathcal{H}$, with kernel κ, which minimizes the regularized empirical risk :

$$\frac{1}{n}\sum_{k=1}^{n} Q(\psi(\boldsymbol{x}_k), y_k) + \eta\, g(\|\psi\|^2_{\mathcal{H}}),$$

which implies n values $\psi(\boldsymbol{x}_k)$ obtained for inputs $\boldsymbol{x}_k$ and (eventually) n desired outputs y_k, with g a strictly monotonic increasing function on $\mathbb{R}^+$, any such function admits a representation of the form :

$$\psi(\cdot) = \sum_{i=1}^{n} \alpha_i \kappa(\cdot, \boldsymbol{x}_i).$$

Sketch of proof :
Any function ψ in $\mathcal{H}$ can be decomposed as $\psi = \sum_{i=1}^{n} \alpha_i\, \kappa(\cdot, \boldsymbol{x}_i) + \psi^{\perp}$, with $\langle \psi^{\perp}, \kappa(\cdot, \boldsymbol{x}_i)\rangle_{\mathcal{H}} = 0$ for all $i = 1, \ldots, n$. Since $\psi(x_j) = \langle \psi, \kappa(\cdot, \boldsymbol{x}_j)\rangle$, the value $\psi(x_j)$ is unaffected by $\psi^{\perp}$, for $j = 1, \ldots, n$.

Consequence :
The minimization on a functional Hilbert space (which can be of infinite dimension) leads to a minimization problem in $\mathbb{R}^n$

Evaluation functional :
A functional (linear) $\delta_{\boldsymbol{x}} : \mathcal{H} \to \mathbb{R}$ is an evaluation functional if it evaluates any function ψ of the space $\mathcal{H}$ at any point $\boldsymbol{x} \in \mathcal{X}$. Thus :

$$\delta_{\boldsymbol{x}}(\psi) = \psi(\boldsymbol{x})$$

Definition (Reproducing Kernel Hilbert Space (RKHS))

A Hilbert space is a *Reproducing Kernel Hilbert Space* if and only if, for any $\boldsymbol{x} \in \mathcal{X}$, the evaluation functional $\delta_{\boldsymbol{x}}$ is bounded.

In other words, there exist M such that for any $\psi \in \mathcal{H}$:

$$|\delta_{\boldsymbol{x}}(\psi)| = |\psi(\boldsymbol{x})| \leq M\|\psi\|_{\mathcal{H}}.$$

Riesz (Fréchet) representation theorem :
If $\mathcal{H}$ is a RKHS, and from the Riesz (Fréchet) representation theorem, for any $\boldsymbol{x} \in \mathcal{X}$ it exists a unique function $\kappa(\cdot, \boldsymbol{x})$ (called representer) from $\mathcal{H}$ such that :

$$\delta_{\boldsymbol{x}}(\psi) = \psi(\boldsymbol{x}) = \langle \psi, \kappa(\cdot, \boldsymbol{x}) \rangle_{\mathcal{H}} \qquad \forall \psi \in \mathcal{H}$$

Reproducing property :

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \kappa(\cdot, \boldsymbol{x}_i), \kappa(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{H}} \qquad\qquad \kappa(\boldsymbol{x}_i, \boldsymbol{x}_i) = \|\kappa(\cdot, \boldsymbol{x}_i)\|^2_{\mathcal{H}}$$

Unicity : For a RKHS, its reproducing Kernel is unique
Sketch of proof :

$$\langle f\left(.\right), \kappa_1\left(., x\right) \rangle = \langle f\left(.\right), \kappa_2\left(., x\right) \rangle$$