MAP 569
Machine Learning II

Christophe Giraud

PC3

RKHS

Reproducing Kernel Hilbert Spaces

Today

1. **Motivations**
2. **RKHS**
3. **Applications**

Motivations

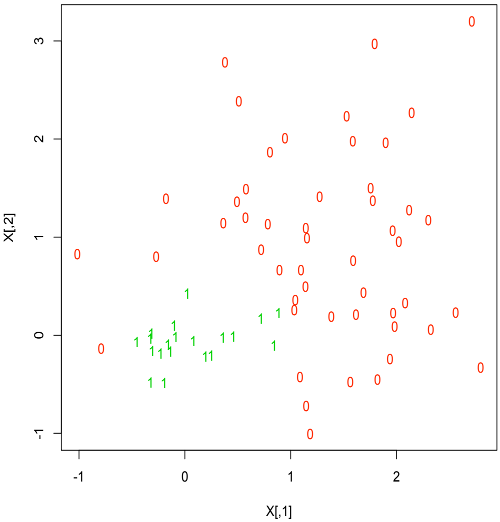
Reproducing Kernel Hilbert Spaces

Goal: to represent the data in a new space $\mathcal{H}$, possibly of infinite dimension.

Why?

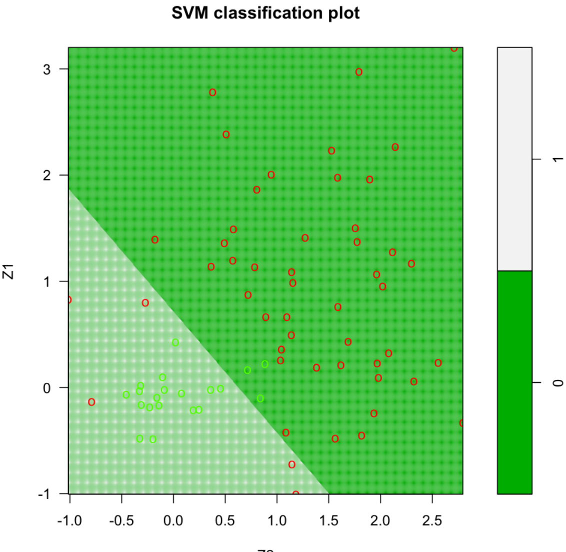
- to "delinearize" an algorithm
- to "vectorialize" data

Example: delinearization 1/4



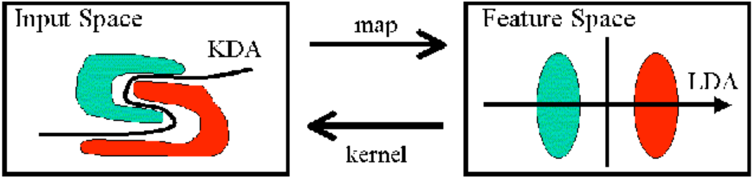
Example: delinearization 2/4

With a linear classifier: poor result



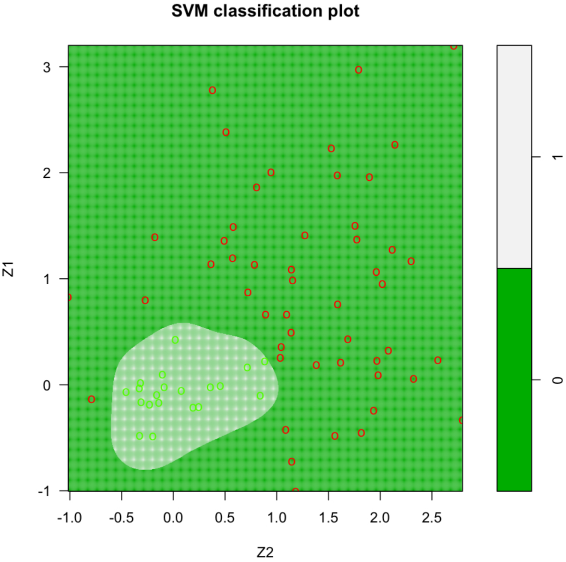
Example: delinearization 3/4

Recipe: apply a linear classifier in a "feature space" $\mathcal{H}$



Example: delinearization 4/4

With a linear classifier in a suitable $\mathcal{H}$

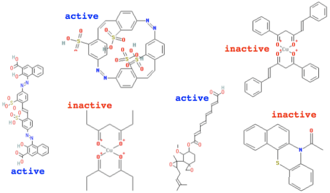


Example: vectorialization of data

How can we handle texts or molecules?

Examples:

- words in email for spam filters. Word = sequence of **variable length**.
- proteins for drugs. Proteins = sequence of **variable length** in the alphabet of 20 amino-acid.
  Insulin: FVNQHLCGSHLVEALYLVCGERGFFYTPKA
- complex molecules for medicine. Molecule = graph.



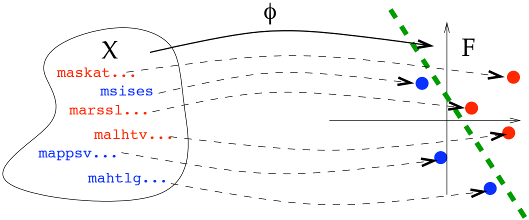
Example: sequences of proteins

Active proteins:

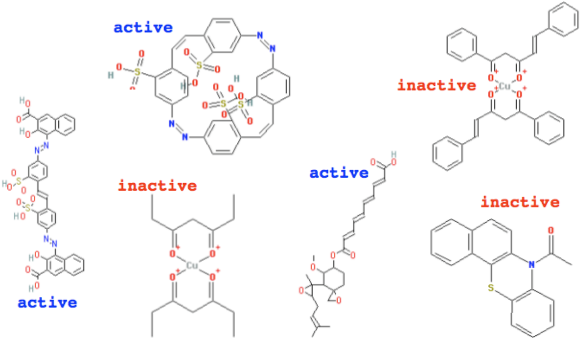
MASKATLLLAFTLLFATCIARHQQRQQQQNQCQLQNIEA...
MARSSLFTFLCLAVFINGCLSQIEQQSPWEFQGSEVW...
MALHTVLIMLSLLPMLEAQNPEHANITIGEPITNETLGWL...
...

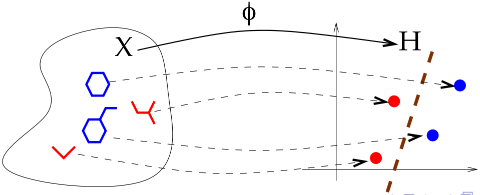
Inactive proteins:

MAPPSVFAEVPQAQPVLVFKLIADFREDPDPRKVNLGVG...
MAHTLGLTQPNSTEPHKISFTAKEIDVIEWKGDILVVG...
MSISESYAKEIKTAFRQFTDFPIEGEQFEDFLPIIGNP.. ...



Example: complex molecules





RKHS

Reproducing kernel

To a function $\phi : \mathcal{X} \to \mathcal{H}$ in an Hilbert space $\mathcal{H}$, we can associate a **"kernel"** $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ defined by

$$k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$$

Property: for all $x_1, \ldots, x_n \in \mathcal{X}$, the matrix $K = [k(x_i, x_j)]_{i,j=1,\ldots,n}$ is positive semi-definite.
Proof?

RKHS: theory 1/3

Let $\mathcal{X}$ be a set.

Notion 1 : positive kernel

A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a **positive kernel** if

- k is symmetric: $k(x, y) = k(y, x)$ for all $x, y \in \mathcal{X}$,
- For all $N \in \mathbb{N}$, $x_1, \ldots, x_N \in \mathcal{X}$ the matrix $K = [k(x_i, x_j)]_{i,j=1,\ldots,n}$ is positive semi-definite.

RKHS: theory 2/3

Notion 2 : RKHS

An Hilbert space $\mathcal{H}$ of functions from $\mathcal{X}$ to $\mathbb{R}$ is called a **Reproducing Kernel Hilbert Space**, if there exists a kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that

1. $k(x, \cdot) \in \mathcal{H}$ for all $x \in \mathcal{X}$
2. Reproduction property: for all $x \in \mathcal{X}$ and $f \in \mathcal{H}$

$$f(x) = \langle f, k(x, \cdot) \rangle_{\mathcal{H}}$$

What is the value of $\langle k(x, .), k(y, .) \rangle_{\mathcal{H}}$?

RKHS: théorie 3/3

Theorem (Aronszajn, 1950)

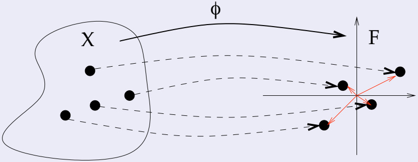
There is a bijection between RKHS and positive kernels

Why?

Geometric picture

The mapping $\phi : \mathcal{X} \to \mathcal{H}$ defined by $\phi(x) = k(x, \cdot)$ fulfills

$$k(x,y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$$



Examples

Examples of kernels: in $\mathbb{R}^d$.

- linear kernel: $k(x, y) = \langle x, y \rangle$
- Gaussian kernel: $k(x, y) = \exp(-|x - y|_2^2 / 2\sigma^2)$
- histogram kernel: $k(x, y) = \min(x, y)$
- exponential kernel: $k(x, y) = \exp(-|x - y|_2 / \sigma)$
- sigmoidal kernel: $k(x, y) = \tanh(a\langle x, y \rangle + b)$
  
  Not positive!!
- etc.

Exercise:

Let us consider the Sobolev space

$$\mathcal{H} = \{f : [0,1] \to \mathbb{R}, \text{ cont., dérivable p.p., } f' \in L^2([0,1]),\ f(0) = 0\}$$

endowed with the Hilbert norm

$$|f|_{\mathcal{H}} = \sqrt{\int_0^1 (f')^2}.$$

1. Prove that if $\mathcal{H}$ is a RKHS with kernel k, then for all $x \in [0,1]$ we have

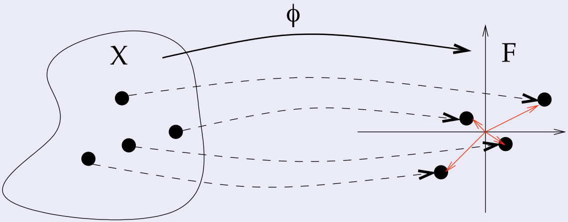
$$f(x) = \int_0^1 f'(y) \frac{\partial}{\partial y} k(x,y)\, dy \quad \text{and} \quad f(x) = \int_0^1 f'(y) \mathbf{1}_{\{y \leq x\}}\, dy.$$

2. What is the kernel k associated to $\mathcal{H}$?
3. What is the shape of $\phi(x) = k(x,.)$?

Geometric picture

The mapping $\phi : \mathcal{X} \to \mathcal{H}$ defined by $\phi(x) = k(x, \cdot)$ fulfills

$$k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$$



Kernel for words/strings

Spectral kernel: basic kernel for words build from an alphabet $\mathcal{A}$.

For $x \in \bigcup_n \mathcal{A}^n$ and $s \in \mathcal{A}^d$ we set

$$N_s(x) = \text{number of occurence of } s \text{ in } x$$

Spectral kernel

For $x, y \in \bigcup_n \mathcal{A}^n$

$$k(x,y) = \sum_{s \in \mathcal{A}^d} N_s(x) N_s(y)$$

Is it positive? how can we compute it efficiently?