MAP 553
Apprentissage statistique

Christophe Giraud

CMAP, Ecole Polytechnique

PC7

Noyaux reproduisants

Au programme:

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

Motivations

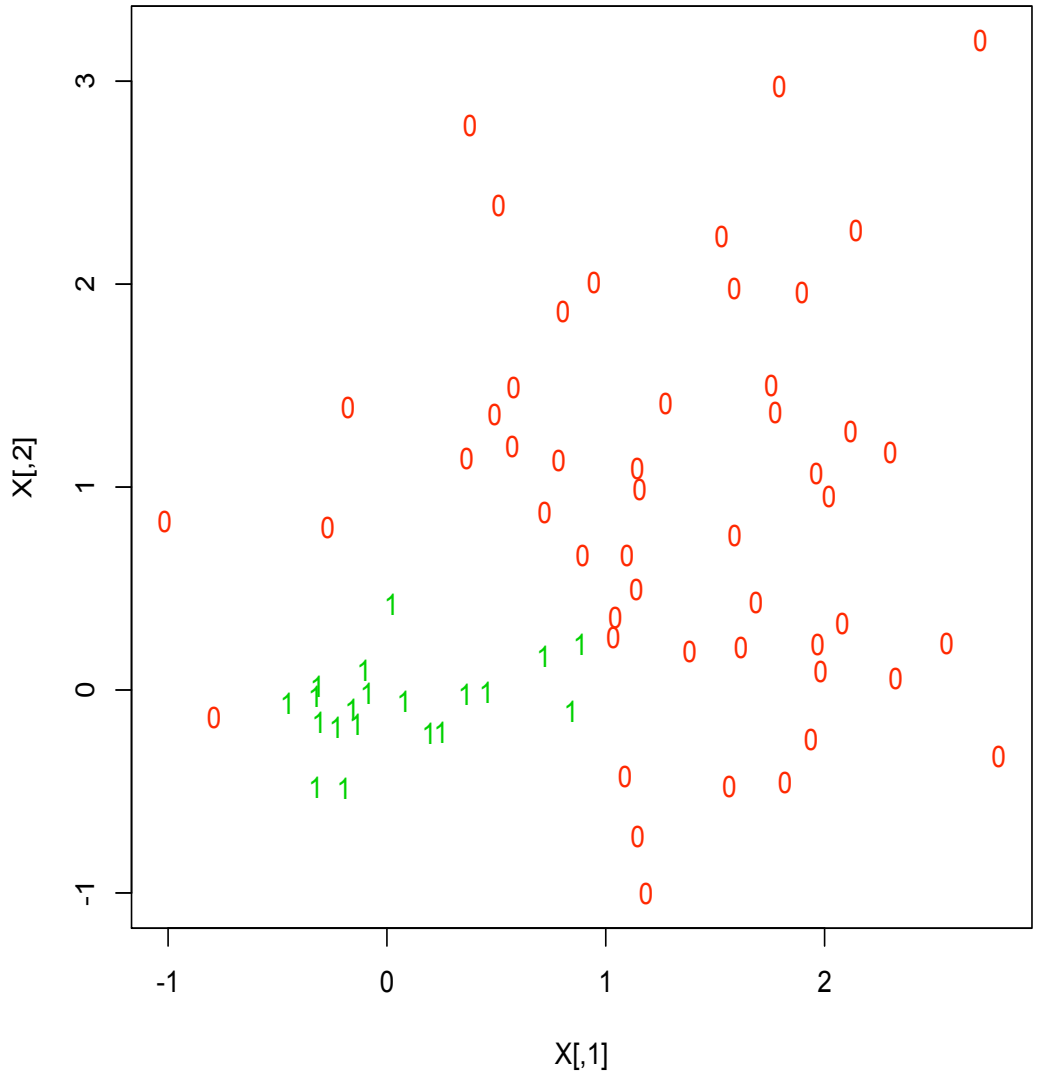
Espaces à noyaux reproduisants

Problème: représenter les données dans un nouvel espace $\mathcal{H}$ de plus grande dimension (souvent infinie).

Dans quel objectif?

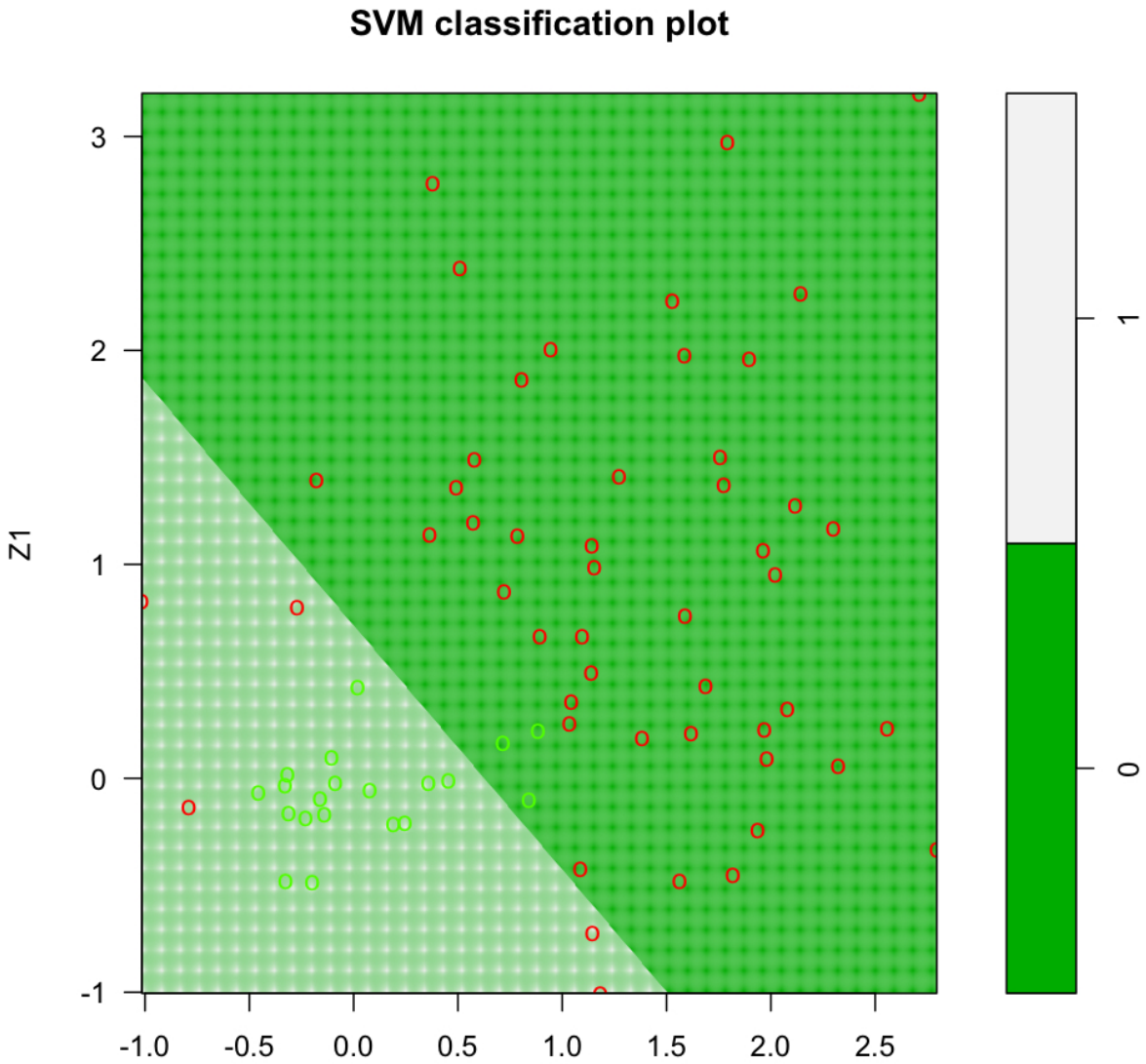
- "délinéariser" un algorithme
- "vectorialiser" les données

Exemple: délinéarisation 1/4



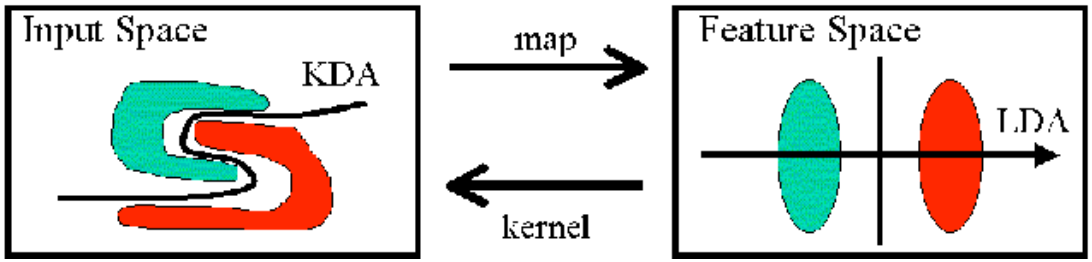
Exemple: délinéarisation 2/4

Avec un classifieur linéaire: résultat médiocre



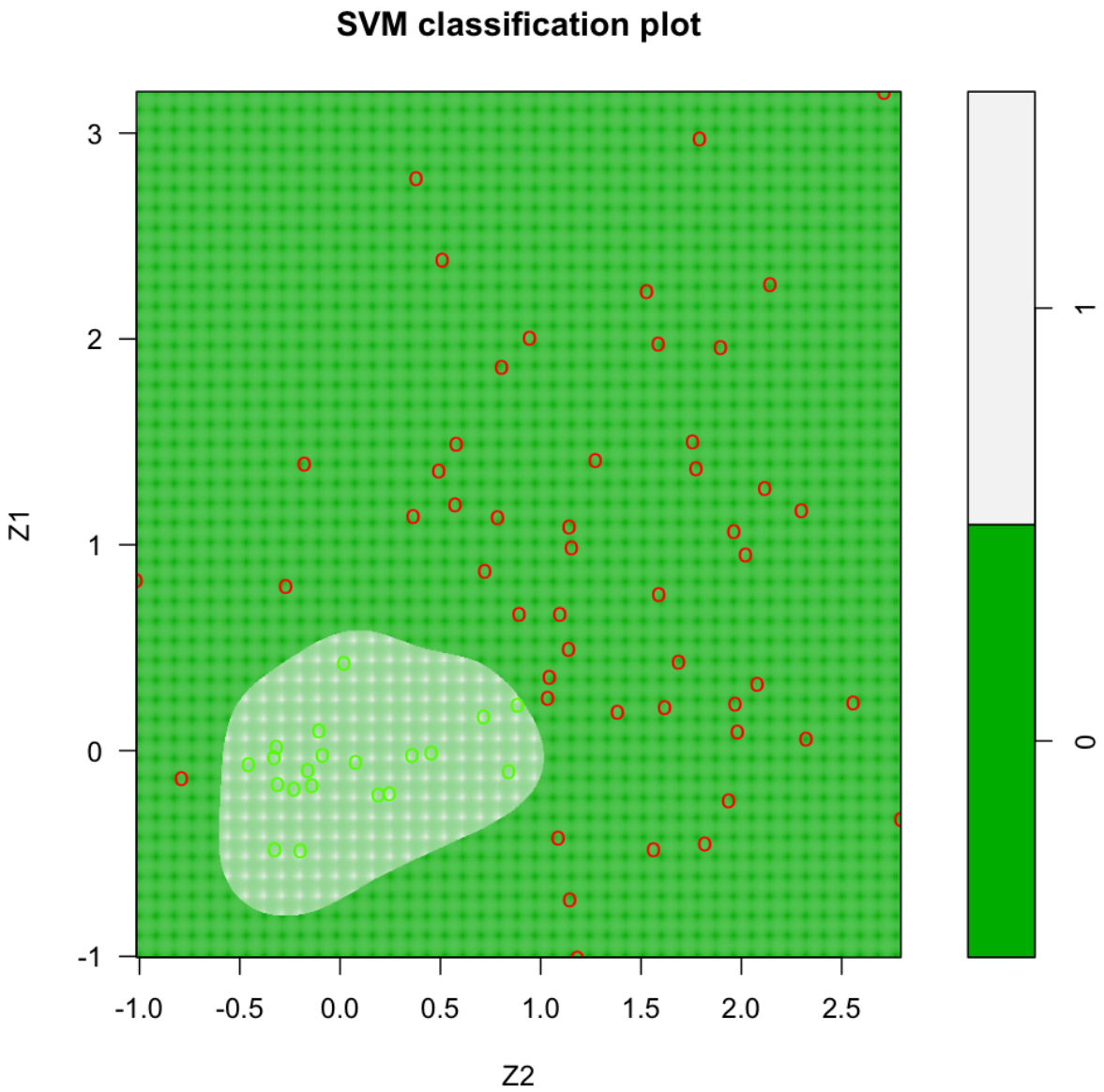
Exemple: délinéarisation 3/4

Idée: utiliser le classifieur linéaire dans un "feature space" $\mathcal{H}$



Exemple: délinéarisation 4/4

Avec le classifieur linéaire dans un espace $\mathcal{H}$ adapté

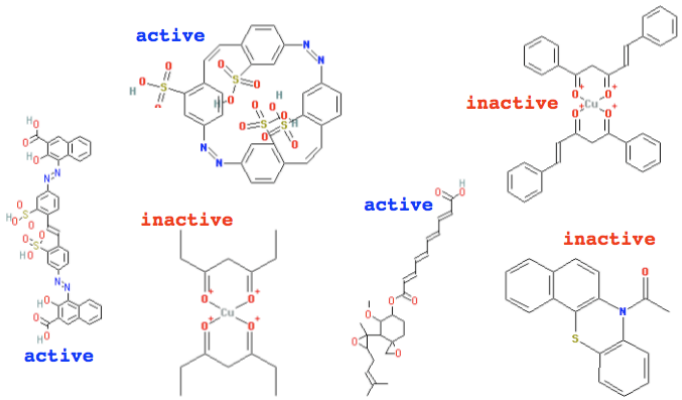


Exemple: vectorialisation des données

Comment travailler avec des textes? des molécules?

Exemples:

- description des mots d'un courriel pour construire un filtre anti-spam. Mot = séquence de **longueur variable**.
- description des protéines pour prédire leur comportement: protéine = séquence de **longueur variable** dans l'alphabet des 20 acides aminés.
  Insuline: FVNQHLCGSHLVEALYLVCGERGFFYTPKA
- description molécules actives / inactives contre HIV pour prédire de potentielles molécules actives. Molécule = graphe.



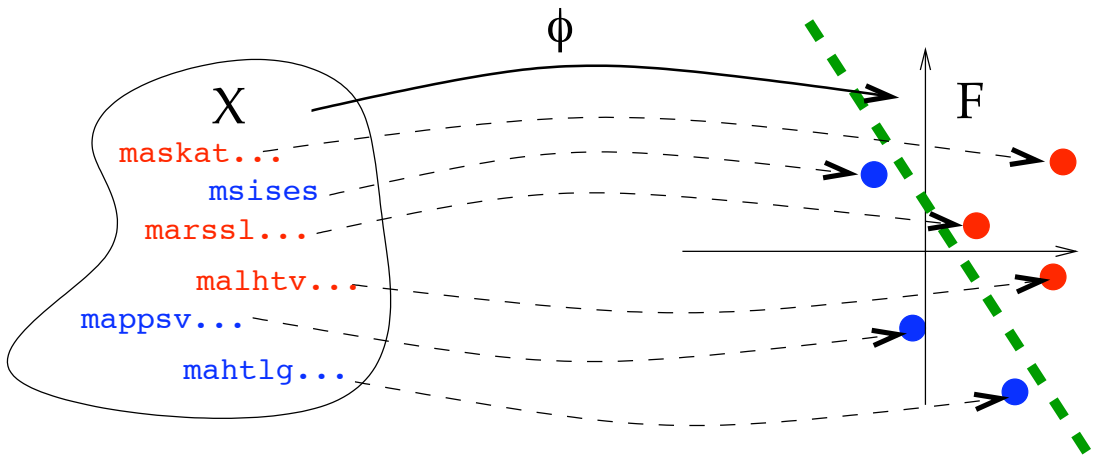
Exemple: séquences de protéines

Protéines sécrétées:

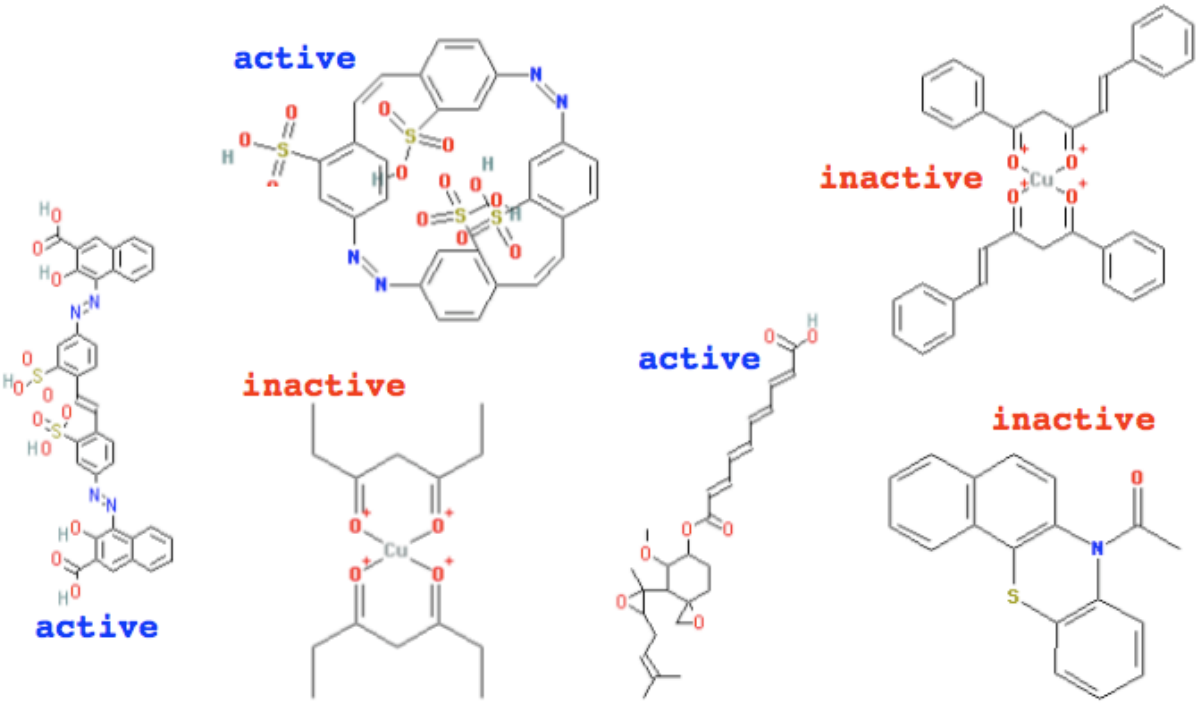
MASKATLLLAFTLLFATCIARHQQRQQQQNQCQLQNIEA...
MARSSLFTFLCLAVFINGCLSQIEQQSPWEFQGSEVW...
MALHTVLIMLSLLPMLEAQNPEHANITIGEPITNETLGWL...
...

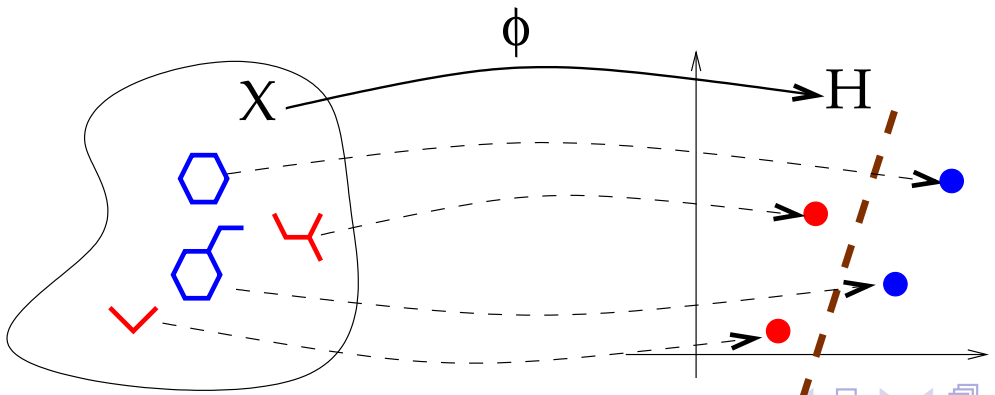
Protéines non-sécrétées:

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



Exemple: molécules médicinales





RKHS: espace de Hilbert à noyau reproduisant

Noyau reproduisant

A une application $\phi : \mathcal{X} \to \mathcal{H}$, avec $\mathcal{H}$ espace de Hilbert, on peut associer le **"noyau"** $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ défini par

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

Propriété: pour $x_1, \ldots, x_n \in \mathcal{X}$, la matrice $K = [k(x_i, x_j)]_{i,j=1,\ldots,n}$ est symétrique et positive.

RKHS: théorie 1/3

Soit $\mathcal{X}$ un ensemble quelconque.

Noyau défini positif

Une fonction $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ est un **noyau défini positif** si

- k est symétrique: $k(x, y) = k(y, x)$ pour tout $x, y \in \mathcal{X}$,
- pour tout $N \in \mathbb{N}$, $x_1, \ldots, x_N \in \mathcal{X}$ et $a_1, \ldots, a_N \in \mathbb{R}$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j k(x_i, x_j) \geq 0.$$

RKHS: théorie 2/3

RKHS

Un espace de Hilbert $\mathcal{H}$ de fonctions sur $\mathcal{X}$ à valeurs réelles est appelé un **espace de Hilbert à noyau reproduisant** ou **RKHS** (en anglais: Reproducing Kernel Hilbert Space), s'il existe un noyau $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ tel que

1. $k(x, \cdot) \in \mathcal{H}$ pour tout $x \in \mathcal{X}$
2. Propriété de reproduction: pour tout $x \in \mathcal{X}$ et $f \in \mathcal{H}$

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

RKHS: théorie 3/3

Théorème (Aronszajn, 1950)

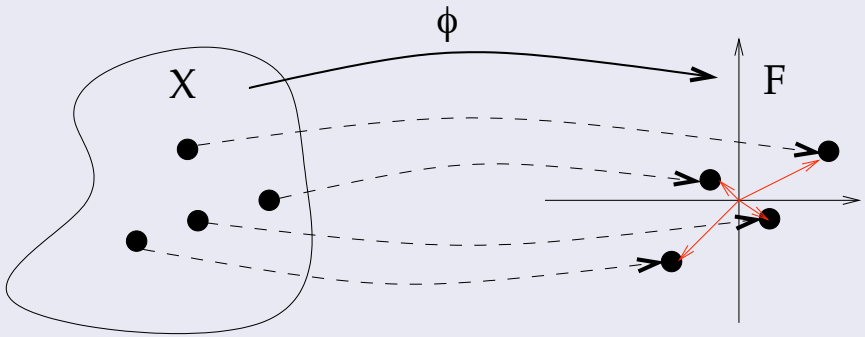
Il existe une bijection entre les RKHS et les noyaux définis positifs.

Admis.

Représentation géométrique

L'application $\phi : \mathcal{X} \to \mathcal{H}$ définie par $\phi(x) = k(x, \cdot)$ vérifie

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



RKHS exemples

Exemples de noyaux: dans $\mathbb{R}^d$.

- noyau linéaire: $k(x, y) = \langle x, y \rangle$
- noyau gaussien: $k(x, y) = \exp(-|x - y|_2^2 / 2\sigma^2)$
- noyau histogramme: $k(x, y) = \min(x, y)$
- noyau exponentiel: $k(x, y) = \exp(-|x - y|_2 / \sigma)$
- noyau sigmoidal: $k(x, y) = \tanh(a\langle x, y \rangle + b)$

 Non défini positif!!

- etc.

Quel est le RKHS associé au noyau linéaire? Gaussien?

Exercice:

Considérons l'espace

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

muni de la norme

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

1. Montrer que pour tout $x \in [0,1]$ on a

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

2. Quel est le noyau reproduisant k associé à $\mathcal{H}$?
3. Quelle forme a la fonction $\phi(x) = k(x, .)$?

Applications

Le "kernel trick"

Tout algorithme qui ne travaille qu'à partir de produits scalaires sur des vecteurs fini-dimensionnels peut être mis en oeuvre dans un RKHS en remplacant les produits scalaires $\langle x, y \rangle$ par $k(x, y)$.

Intérêt:

- fondammental en pratique!
- pas besoin de connaitre le RKHS, seul k importe pour les calculs (les vecteurs dans $\mathcal{H}$ sont manipulés implicitement).

Exemples:

- ACP à noyau
- ALD à noyau
- SVM à noyau
- etc. (la liste est longue)

ACP à Noyau (1/3)

Soit $x_1, \ldots, x_n \in \mathcal{X}$ et $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ un noyau. On veut faire une ACP dans $\mathcal{H}$ à partir des observations $\phi(x_1), \ldots, \phi(x_n)$.

Quitte à remplacer $\phi(x_j)$ par $\phi(x_j) - \frac{1}{n}\sum_i \phi(x_i)$ pour $j = 1, \ldots, n$, on supposera que $\frac{1}{n}\sum_i \phi(x_i) = 0$. Cela revient à remplacer $K = [k(x_i, x_j)]_{i,j=1,\ldots,n}$ par $K - U_n K - K U_n + U_n K U_n$ où $U_n = [1/n]_{i,j=1,\ldots,n}$.

On notera $b_1, \ldots, b_n$ des vecteurs propres orthonormés, associés aux valeurs propres $\lambda_1 \geq \ldots \geq \lambda_n$ de K.

1. Montrer que
$$\max_{|f|^2_{\mathcal{H}} \leq 1} \sum_{i=1}^{n} \langle \phi(x_i), f \rangle^2_{\mathcal{H}} \;=\; \max_{\alpha^T K \alpha \leq 1} \alpha^T K^2 \alpha.$$
2. En déduire que le premier axe est donné par
$$f^{(1)} = \sum_{i=1}^{n} \alpha_i^{(1)} \phi(x_i) \quad \text{où} \quad \alpha^{(1)} = b_1/\sqrt{\lambda_1}$$

ACP à Noyau (2/3)

3. Montrez que les fonctions $f = \sum_i \alpha_i \phi(x_i)$ et $g = \sum_i \beta_i \phi(x_i)$ sont orthogonales si et seulement si $\alpha^T K \beta = 0$. En déduire que le k-ième axe est donné par

$$f^{(k)} = \sum_{i=1}^{n} \alpha_i^{(k)} \phi(x_i) \quad \text{où} \quad \alpha^{(k)} = b_k / \sqrt{\lambda_k}$$

4. Montrez que pour tout $x \in \mathcal{X}$, la coordonnée de $\phi(x)$ sur l'axe $f^{(k)}$ est donnée par

$$\langle \phi(x), f^{(k)} \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \alpha_i^{(k)} k(x_i, x).$$

En particulier, $\langle \phi(x_i), f^{(k)} \rangle_{\mathcal{H}} = [K\alpha^{(k)}]_i$ pour $i = 1, \ldots, n$.

ACP à Noyau (3/3)

Synthèse

Pour réaliser une ACP à noyau avec q axes:

1. remplacer K par $K - U_n K - K U_n + U_n K U_n$
2. calculer $(b_k, \lambda_k)_{k=1,\ldots,q}$
3. la projection du point $\phi(x_i)$ sur les q premiers axes est représentée par le vecteur $[(K\alpha^{(1)})_i, \ldots, (K\alpha^{(q)})_i]$.

Exemple: noyau spectral

Noyau spectral: noyau générique pour représenter des "mots" ou séquences de "mots" écrits à partir d'un alphabet $\mathcal{A}$.

Pour $x \in \bigcup_n \mathcal{A}^n$ et $s \in \mathcal{A}^k$ on note

$$N_s(x) = \text{nombre d'occurences de } s \text{ dans } x$$

Noyau spectral

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

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

est-il défini positif? temps de calcul de $k(x,y)$?

Pour en savoir plus...

Jean-Philippe Vert

Mines ParisTech
Centre for Computational Biology

Institut Curie
Bioinformatics and Computational
Systems Biology of Cancer

Slides de son cours au master MVA:

`http://cbio.ensmp.fr/~jvert/teaching/2010mva/index.html`

Autre référence:

B. Schölkopf and A.J. Smola. *Learning with Kernels.* MIT Press.