Méthodes à Noyaux pour L'analyse de Données

Mathieu Fauvel

gipsa-lab

19 janvier 2007

Représentation des données :

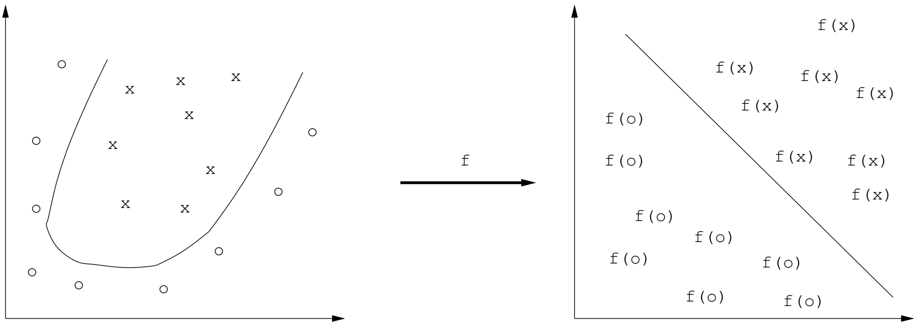
- $\mathbf{x} \in \mathbb{R}$ ou $\mathbb{C}$
- $\mathbf{x} \in \mathbb{R}^n$ ou $\mathbb{C}^n$
- $\mathbf{x} \in \mathbb{R}^n \times \mathbb{R}^n$ ou $\mathbb{C}^n \times \mathbb{C}^n$
- $\mathbf{x} \in (\Omega, \sigma, P)$
- $\mathbf{x} \in$ ensemble de toutes les séquences finies de l'alphabet $\{A, C, G, T\}$ ($\mathbf{x}$ est une séquences d'ADN)
- $\mathbf{x} \subset \mathcal{P}(E)$
- $\ldots$

Notations : $\mathcal{X}$ un ensemble quelconque, un objet $\mathbf{x} \in \mathcal{X}$ et $\mathcal{S} = \{\mathbf{x}^1, \ldots, \mathbf{x}^N\}$ un ensemble fini de N objets à notre disposition.

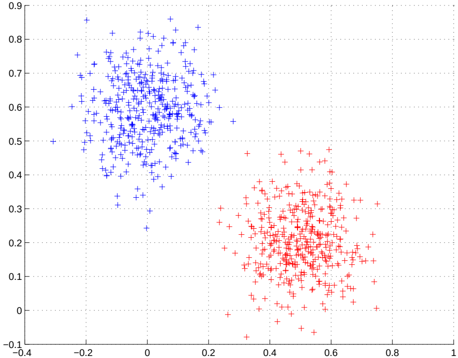
Méthodes à noyaux : Transformer tout algorithme linéaire, s'exprimant sous forme de produits scalaires, en un algorithme non linéaire.

Idée de Base :

- Projeter les données dans un espace vectoriel
- Chercher des relations linéaires dans cet espace



Classification : $\mathcal{S} = \{(\mathbf{x}^i, y_i) \in \mathbb{R}^n \times \{-1; 1\};\ i \in [1, N]\}$



$$\|M_1 - \mathbf{x}\|^2 \overset{?}{\gtrless} \|M_{-1} - \mathbf{x}\|^2$$

$$\|M_1 - \mathbf{x}\|^2 - \|M_{-1} - \mathbf{x}\|^2 \overset{?}{\gtrless} 0$$

$$\begin{aligned}
\|M_1 - \mathbf{x}\|^2 &= \left\langle \frac{1}{m_1}\sum_{i=1}^{m_1}\mathbf{x}^i - \mathbf{x}, \frac{1}{m_1}\sum_{i=1}^{m_1}\mathbf{x}^i - \mathbf{x} \right\rangle \\
&= \left\langle \frac{1}{m_1}\sum_{i=1}^{m_1}\mathbf{x}^i, \frac{1}{m_1}\sum_{i=1}^{m_1}\mathbf{x}^i \right\rangle + \langle \mathbf{x}, \mathbf{x}\rangle - 2\left\langle \frac{1}{m_1}\sum_{i=1}^{m_1}\mathbf{x}^i, \mathbf{x} \right\rangle \\
&= \frac{1}{m_1^2}\sum_{\substack{i=1\\k=1}}^{m_1}\left\langle \mathbf{x}^i, \mathbf{x}^k \right\rangle + \langle \mathbf{x}, \mathbf{x}\rangle - 2\left\langle \frac{1}{m_1}\sum_{i=1}^{m_1}\mathbf{x}^i, \mathbf{x} \right\rangle \\
\|M_{-1} - \mathbf{x}\|^2 &= \frac{1}{m_{-1}^2}\sum_{\substack{j=1\\l=1}}^{m_{-1}}\left\langle \mathbf{x}^j, \mathbf{x}^l \right\rangle + \langle \mathbf{x}, \mathbf{x}\rangle - 2\left\langle \frac{1}{m_{-1}}\sum_{j=1}^{m_{-1}}\mathbf{x}^j, \mathbf{x} \right\rangle
\end{aligned}$$

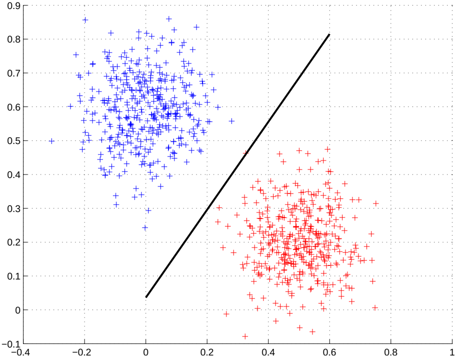
Règle de décision :

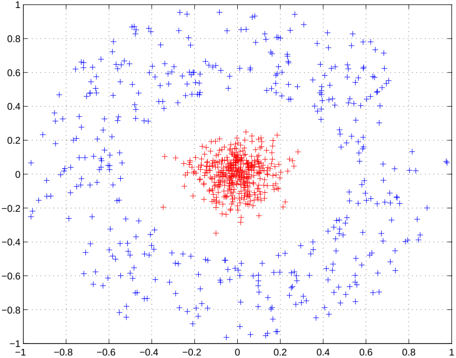
$$\left\langle -2\left(\frac{1}{m_1}\sum_{i=1}^{m_1}\mathbf{x}^i - \frac{1}{m_{-1}}\sum_{j=1}^{m_{-1}}\mathbf{x}^j\right), \mathbf{x}\right\rangle + \frac{1}{m_1^2}\sum_{\substack{i=1\\k=1}}^{m_1}\left\langle \mathbf{x}^i, \mathbf{x}^k\right\rangle - \frac{1}{m_{-1}^2}\sum_{\substack{j=1\\l=1}}^{m_{-1}}\left\langle \mathbf{x}^j, \mathbf{x}^l\right\rangle \overset{?}{\gtrless} 0$$

Étudier le signe :

$$\begin{aligned}
f(\mathbf{x}) &= \langle \mathbf{w}, \mathbf{x}\rangle + b \\
\mathbf{w} &= -2\sum_{m=1}^{m_1+m_{-1}} \alpha_m y_m \mathbf{x}^m \\
\alpha_m &= \frac{1}{m_1} \text{ ou } \frac{1}{m_{-1}} \\
y_m &= 1 \text{ ou } -1 \\
b &= \frac{1}{m_1^2}\sum_{\substack{i=1\\k=1}}^{m_1}\left\langle \mathbf{x}^i, \mathbf{x}^k\right\rangle - \frac{1}{m_{-1}^2}\sum_{\substack{j=1\\l=1}}^{m_{-1}}\left\langle \mathbf{x}^j, \mathbf{x}^l\right\rangle
\end{aligned}$$

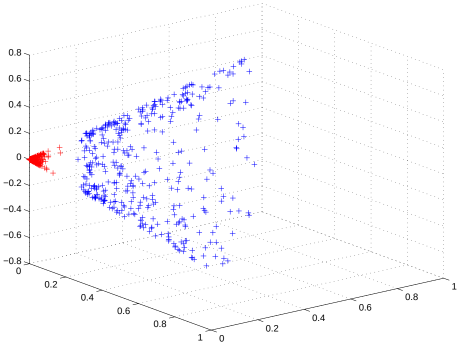
Séparation : $\{\mathbf{x}|f(\mathbf{x}) = 0\}$.

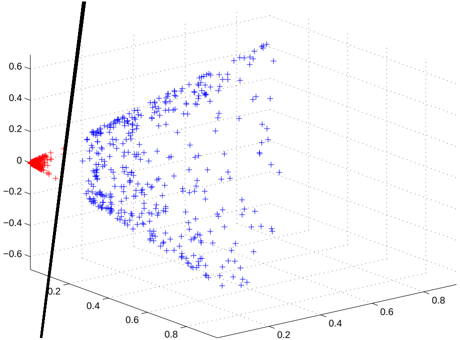




Projection :

$$\begin{aligned} \Phi : \mathbb{R}^2 &\rightarrow \mathbb{R}^3 \\ \mathbf{x} &\mapsto \Phi(\mathbf{x}) \\ (x_1, x_2) &\mapsto (x_1^2, x_2^2, \sqrt{2}x_1x_2) \end{aligned}$$


0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
0
0.2
0.4
0.6
0.8
1
0
0.2
0.4
0.6
0.8
1

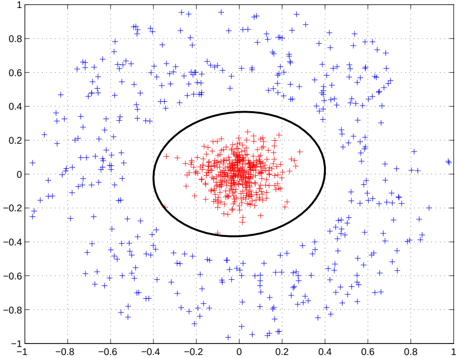

0.6
0.4
0.2
0
−0.2
−0.4
−0.6
0.2
0.4
0.6
0.8
0.2
0.4
0.6
0.8

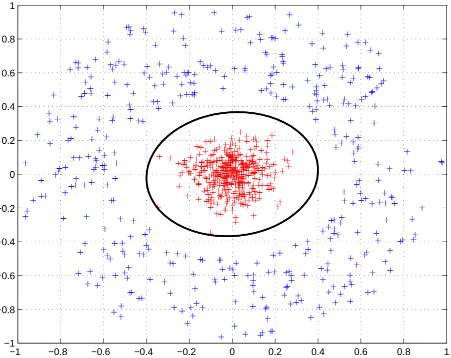
Ré-écriture du produit scalaire : $\langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathbb{R}^3}$

$$\begin{aligned}
\sum_{i=1}^{3} \Phi(\mathbf{x})_i \Phi(\mathbf{y})_i &= \Phi(\mathbf{x})_1 \Phi(\mathbf{y})_1 + \Phi(\mathbf{x})_2 \Phi(\mathbf{y})_2 + \Phi(\mathbf{x})_3 \Phi(\mathbf{y})_3 \\
&= {x_1}^2 {y_1}^2 + {x_2}^2 {y_2}^2 + 2 x_1 x_2 y_1 y_2 \\
&= (x_1 y_1 + x_2 y_2)^2 \\
&= \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}^2}^2 \\
&= k(\mathbf{x}, \mathbf{y})
\end{aligned}$$

Ré-écriture de la règle de décision :

$$\begin{aligned}
f(\mathbf{x}) &= k(\mathbf{w}, \mathbf{x}) + b \\
b &= \frac{1}{m_1^2} \sum_{\substack{i=1 \\ k=1}}^{m_1} k(\mathbf{x}^i, \mathbf{x}^k) - \frac{1}{m_{-1}^2} \sum_{\substack{j=1 \\ l=1}}^{m_{-1}} k(\mathbf{x}^j, \mathbf{x}^l)
\end{aligned}$$


1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1



Pour quelles fonctions k a-t on :

$$k(\mathbf{x}, \mathbf{y}) \quad = \quad \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathcal{H}}$$

Pour quelles fonctions k a-t-on l'égalité suivante :

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

avec

$$\begin{aligned} k : \mathcal{X} \times \mathcal{X} &\rightarrow \mathbb{R} \\ (\mathbf{x}, \mathbf{y}) &\mapsto k(\mathbf{x}, \mathbf{y}) \end{aligned} \qquad \begin{aligned} \Phi : \mathcal{X} &\rightarrow \mathcal{H} \\ \mathbf{x} &\mapsto \Phi(\mathbf{x}) \end{aligned}$$

Theorem (Théorème de Moore-Aronsjan (1950))

A chaque noyau semi-défini positif $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, on peut associer un unique espace de Hilbert $(\mathcal{H}, \langle , \rangle_{\mathcal{H}})$ ayant k comme noyau reproduisant :

$$\forall f \in \mathcal{H}, \forall : \mathbf{x} \in \mathcal{X},\ f(\mathbf{x}) = \langle f, k(., \mathbf{x}) \rangle_{\mathcal{H}}$$

Noyau semi défini positif : $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ est semi défini positif si :

1. $\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{X};\ k(\mathbf{x}, \mathbf{y}) = k(\mathbf{y}, \mathbf{x})$
2. $\forall n \in \mathbb{N}, \forall \xi_1 \dots \xi_n \in \mathbb{R}, \forall \mathbf{x}^1 \dots \mathbf{x}^n \in \mathcal{X}, \sum_{i,j}^{n} \xi_i \xi_j k(\mathbf{x}^i, \mathbf{x}^j) \geq 0$

Espace de Banach : Espace vectoriel normé complet pour la distance issue de sa norme.

Espace de Hilbert : Espace de Banach dont la norme $\|.\|$ découle d'un produit scalaire : $\|\mathbf{x}\|_{\mathcal{H}} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{H}}}$.

Espace de Hilbert à noyau auto-reproduisant : Soit $\mathcal{X}$ un espace quelconque, $\mathcal{H}$ un espace vectoriel de fonctions définies sur $\mathcal{X}$ à valeurs réelles ; $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ et $(\mathcal{H}, \langle , \rangle_{\mathcal{H}})$ un espace de Hilbert. La fonction $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ est un noyau reproduisant si :

1. $\forall \mathbf{y} \in \mathcal{X} :\ K(., \mathbf{y}) : \mathbf{x} \mapsto K(\mathbf{x}, \mathbf{y}) \in \mathcal{H}$
2. $\forall \mathbf{y} \in \mathcal{X}$ et $\forall f \in \mathcal{H} : f(\mathbf{y}) = \langle f, K(., \mathbf{y}) \rangle_{\mathcal{H}}$

Propriété reproduisante : $K(\mathbf{x}, \mathbf{y}) = \langle K(., \mathbf{x}), K(., \mathbf{y}) \rangle_{\mathcal{H}}$.

Noyau semi défini positif : $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ est semi défini positif si :

1. $\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{X};\ k(\mathbf{x}, \mathbf{y}) = k(\mathbf{y}, \mathbf{x})$
2. $\forall n \in \mathbb{N}, \forall \xi_1 \dots \xi_n \in \mathbb{R}, \forall \mathbf{x}^1 \dots \mathbf{x}^n \in \mathcal{X}, \sum_{i,j}^{n} \xi_i \xi_j k(\mathbf{x}^i, \mathbf{x}^j) \geq 0$

Espace de Banach : Espace vectoriel normé complet pour la distance issue de sa norme.

Espace de Hilbert : Espace de Banach dont la norme $\|.\|$ découle d'un produit scalaire : $\|\mathbf{x}\|_{\mathcal{H}} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{H}}}$.

Espace de Hilbert à noyau auto-reproduisant : Soit $\mathcal{X}$ un espace quelconque, $\mathcal{H}$ un espace vectoriel de fonctions définies sur $\mathcal{X}$ à valeurs réelles ; $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ et $(\mathcal{H}, \langle , \rangle_{\mathcal{H}})$ un espace de Hilbert. La fonction $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ est un noyau reproduisant si :

1. $\forall \mathbf{y} \in \mathcal{X} :\ K(., \mathbf{y}) : \mathbf{x} \mapsto K(\mathbf{x}, \mathbf{y}) \in \mathcal{H}$
2. $\forall \mathbf{y} \in \mathcal{X}$ et $\forall f \in \mathcal{H} : f(\mathbf{y}) = \langle f, K(., \mathbf{y}) \rangle_{\mathcal{H}}$

Propriété reproduisante : $K(\mathbf{x}, \mathbf{y}) = \langle K(., \mathbf{x}), K(., \mathbf{y}) \rangle_{\mathcal{H}}$.

Noyau semi défini positif : $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ est semi défini positif si :

1. $\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{X};\ k(\mathbf{x}, \mathbf{y}) = k(\mathbf{y}, \mathbf{x})$
2. $\forall n \in \mathbb{N}, \forall \xi_1 \dots \xi_n \in \mathbb{R}, \forall \mathbf{x}^1 \dots \mathbf{x}^n \in \mathcal{X}, \sum_{i,j}^n \xi_i \xi_j k(\mathbf{x}^i, \mathbf{x}^j) \geq 0$

Espace de Banach : Espace vectoriel normé complet pour la distance issue de sa norme.

Espace de Hilbert : Espace de Banach dont la norme $\|.\|$ découle d'un produit scalaire : $\|\mathbf{x}\|_{\mathcal{H}} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{H}}}$.

Espace de Hilbert à noyau auto-reproduisant : Soit $\mathcal{X}$ un espace quelconque, $\mathcal{H}$ un espace vectoriel de fonctions définies sur $\mathcal{X}$ à valeurs réelles ; $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ et $(\mathcal{H}, \langle , \rangle_{\mathcal{H}})$ un espace de Hilbert. La fonction $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ est un noyau reproduisant si :

1. $\forall \mathbf{y} \in \mathcal{X} :\ K(., \mathbf{y}) : \mathbf{x} \mapsto K(\mathbf{x}, \mathbf{y}) \in \mathcal{H}$
2. $\forall \mathbf{y} \in \mathcal{X}$ et $\forall f \in \mathcal{H} : f(\mathbf{y}) = \langle f, K(., \mathbf{y}) \rangle_{\mathcal{H}}$

Propriété reproduisante : $K(\mathbf{x}, \mathbf{y}) = \langle K(., \mathbf{x}), K(., \mathbf{y}) \rangle_{\mathcal{H}}$.

Propriétés des noyaux auto-reproduisant (a.r.) :

1. Si K existe, alors il est unique
2. Un noyau a.r. est un noyau semi-défini positif :

Preuve 2.

- $K(\mathbf{x},\mathbf{y}) = \langle K(.,\mathbf{x}), K(.,\mathbf{y})\rangle_{\mathcal{H}} = \langle K(.,\mathbf{y}), K(.,\mathbf{x})\rangle_{\mathcal{H}} = K(\mathbf{y},\mathbf{x})$
- $\sum_{\substack{i=1\\j=1}}^{n} \xi_i\xi_j K(\mathbf{x}^i,\mathbf{x}^j) = \sum_{\substack{i=1\\j=1}}^{n} \xi_i\xi_j \langle K(.,\mathbf{x}^i), K(.,\mathbf{x}^j)\rangle_{\mathcal{H}} =$

$$\left\langle \sum_{i=1}^{n} \xi_i K(.,\mathbf{x}^i), \sum_{j=1}^{n} \xi_j K(.,\mathbf{x}^j) \right\rangle_{\mathcal{H}} = \left\| \sum_{i=1}^{n} \xi_i K(.,\mathbf{x}^i) \right\|_{\mathcal{H}}^2 \geq 0$$

$\square$

Construction de l'espace de Hilbert :
k : un noyau semi-défini positif, $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$
$\mathcal{H}_0$ espace vectoriel des combinaisons linéaires finies :

$$f := \sum_{i=1}^{p} \alpha_i k(., \mathbf{x}^i) | \mathbf{x}^i \in \mathcal{X},\ \alpha_i \in \mathbb{R} \text{ et } p \in \mathbb{N}$$

On définit :

$$\langle f, g \rangle_{\mathcal{H}_0} := \sum_{i,j=1}^{p,q} \alpha_i \beta_j k(\mathbf{x}^i, \mathbf{y}^j)$$

Montrons que $\langle f, g \rangle_{\mathcal{H}_0}$ est un produit scalaire :

- Indépendant de la base de représentation
- bilinéaire (*immédiat*)
- symétrique (*immédiat*)
- positif
- défini

Indépendance de la représentation de f et g.

$$\langle f, g\rangle_{\mathcal{H}_0} = \sum_{i=1}^{p}\alpha_i \sum_{j=1}^{q}\beta_j k(\mathbf{x}^i, \mathbf{y}^j) = \sum_{i=1}^{p}\alpha_i g(\mathbf{x}^i) = \sum_{j=1}^{q}\beta_j f(\mathbf{y}^j)$$ □

Positivité.

$\langle f, f\rangle_{\mathcal{H}_0} = \sum_{\substack{i=1\\j=1}}^{p}\alpha_i\alpha_j k(\mathbf{x}^i, \mathbf{x}^j) \geq 0, \forall \alpha_i, \mathbf{x}^i, p$. k est un noyau semi-défini positif. □

Défini.

Remarquons : $\langle f, k(., \mathbf{x})\rangle_{\mathcal{H}_0} = \sum_{i=1}^{p}\alpha_i k(\mathbf{x}^i, \mathbf{x}) = f(\mathbf{x})$ (*prop. reproduisante*).

$|f(\mathbf{x})| = |\langle f, k(, \mathbf{x})\rangle_{\mathcal{H}_0}| \leq \|f\|_{\mathcal{H}_0} k(\mathbf{x}, \mathbf{x})^{1/2}$ (*bilinéarité + positivité + symétrie*)

Donc $\|f\|_{\mathcal{H}_0} = 0 \Rightarrow (f(\mathbf{x}) = 0, \forall \mathbf{x}) \Rightarrow f = 0$ □

Indépendance de la représentation de f et g.

$$\langle f, g\rangle_{\mathcal{H}_0} = \sum_{i=1}^{p} \alpha_i \sum_{j=1}^{q} \beta_j k(\mathbf{x}^i, \mathbf{y}^j) = \sum_{i=1}^{p} \alpha_i g(\mathbf{x}^i) = \sum_{j=1}^{q} \beta_j f(\mathbf{y}^j)$$

□

Positivité.

$\langle f, f\rangle_{\mathcal{H}_0} = \sum_{\substack{i=1\\j=1}}^{p} \alpha_i \alpha_j k(\mathbf{x}^i, \mathbf{x}^j) \geq 0, \forall \alpha_i, \mathbf{x}^i, p$. k est un noyau semi-défini positif. □

Défini.

Remarquons : $\langle f, k(., \mathbf{x})\rangle_{\mathcal{H}_0} = \sum_{i=1}^{p} \alpha_i k(\mathbf{x}^i, \mathbf{x}) = f(\mathbf{x})$ (*prop. reproduisante*).

$|f(\mathbf{x})| = |\langle f, k(, \mathbf{x})\rangle_{\mathcal{H}_0}| \leq \|f\|_{\mathcal{H}_0} k(\mathbf{x}, \mathbf{x})^{1/2}$ (*bilinéarité + positivité + symétrie*)

Donc $\|f\|_{\mathcal{H}_0} = 0 \Rightarrow (f(\mathbf{x}) = 0, \forall \mathbf{x}) \Rightarrow f = 0$ □

Indépendance de la représentation de f et g.

$$\langle f, g\rangle_{\mathcal{H}_0} = \sum_{i=1}^{p} \alpha_i \sum_{j=1}^{q} \beta_j k(\mathbf{x}^i, \mathbf{y}^j) = \sum_{i=1}^{p} \alpha_i g(\mathbf{x}^i) = \sum_{j=1}^{q} \beta_j f(\mathbf{y}^j)$$ $\square$

Positivité.

$\langle f, f\rangle_{\mathcal{H}_0} = \sum_{\substack{i=1\\j=1}}^{p} \alpha_i\alpha_j k(\mathbf{x}^i, \mathbf{x}^j) \geq 0, \forall \alpha_i, \mathbf{x}^i, p$. k est un noyau semi-défini positif. $\square$

Défini.

Remarquons : $\langle f, k(., \mathbf{x})\rangle_{\mathcal{H}_0} = \sum_{i=1}^{p} \alpha_i k(\mathbf{x}^i, \mathbf{x}) = f(\mathbf{x})$ (*prop. reproduisante*).

$|f(\mathbf{x})| = |\langle f, k(, \mathbf{x})\rangle_{\mathcal{H}_0}| \leq \|f\|_{\mathcal{H}_0} k(\mathbf{x}, \mathbf{x})^{1/2}$ (*bilinéarité + positivité + symétrie*)

Donc $\|f\|_{\mathcal{H}_0} = 0 \Rightarrow (f(\mathbf{x}) = 0, \forall \mathbf{x}) \Rightarrow f = 0$ $\square$

Theorem (Complété)

Si $(\mathcal{E}_0, \langle , \rangle_{\mathcal{E}_0})$ *est un espace métrique, il existe un espace métrique* $(\widetilde{\mathcal{E}_0}, \langle , \rangle_{\widetilde{\mathcal{E}_0}})$ *complet dont* $(\mathcal{E}_0, \langle , \rangle_{\mathcal{E}_0})$ *est un sous espace dense. Cet espace est unique à isométrie près. On l'appelle le complété de* $(\mathcal{E}_0, \langle , \rangle_{\mathcal{E}_0})$

$(\widetilde{\mathcal{H}_0}, \langle , \rangle_{\widetilde{\mathcal{H}_0}})$, complété de $(\mathcal{H}_0, \langle , \rangle_{\mathcal{H}_0})$, est un espace *Hilbertien*. ■

Conclusion :

k semi-defini positif $\iff k(\mathbf{x}, \mathbf{y}) = \langle k(., \mathbf{x}), k(., \mathbf{y}) \rangle_{\mathcal{H}} = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathcal{H}}$

Theorem (Complété)

Si $(\mathcal{E}_0, \langle , \rangle_{\mathcal{E}_0})$ *est un espace métrique, il existe un espace métrique* $(\widetilde{\mathcal{E}_0}, \langle , \rangle_{\widetilde{\mathcal{E}_0}})$ *complet dont* $(\mathcal{E}_0, \langle , \rangle_{\mathcal{E}_0})$ *est un sous espace dense. Cet espace est unique à isométrie près. On l'appelle le complété de* $(\mathcal{E}_0, \langle , \rangle_{\mathcal{E}_0})$

$(\widetilde{\mathcal{H}_0}, \langle , \rangle_{\widetilde{\mathcal{H}_0}})$, complété de $(\mathcal{H}_0, \langle , \rangle_{\mathcal{H}_0})$, est un espace *Hilbertien*. ■

Conclusion :

$$k \text{ semi-defini positif} \iff k(\mathbf{x}, \mathbf{y}) = \langle k(., \mathbf{x}), k(., \mathbf{y}) \rangle_{\mathcal{H}} = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathcal{H}}$$

$(\mathcal{X}, \langle , \rangle_{\mathcal{X}})$ espace de Hilbert $\Rightarrow k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{X}}$ est semi-défini positif.

- $k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{X}} = \langle \mathbf{y}, \mathbf{x} \rangle_{\mathcal{X}} = k(\mathbf{y}, \mathbf{x})$
- $\displaystyle\sum_{\substack{i=1\\j=1}}^{n} \xi_i \xi_j \langle \mathbf{x}^i, \mathbf{x}^j \rangle_{\mathcal{X}} = \left\langle \sum_{i=1}^{n} \xi_i \mathbf{x}^i, \sum_{j=1}^{n} \xi_j \mathbf{x}^j \right\rangle_{\mathcal{X}} = \left\| \sum_{i=1}^{n} \xi_i \mathbf{x}^i \right\|_{\mathcal{X}}^2 \geq 0$

$\square$

$g : \mathcal{X} \to \mathbb{R} \Rightarrow k(\mathbf{x}, \mathbf{y}) = g(\mathbf{x}) \cdot g(\mathbf{y})$ est semi-défini positif.

- $k(\mathbf{x}, \mathbf{y}) = g(\mathbf{x}) \cdot g(\mathbf{y}) = g(\mathbf{y}) \cdot g(\mathbf{x}) = k(\mathbf{y}, \mathbf{x})$
- $\displaystyle\sum_{\substack{i=1\\j=1}}^{n} \xi_i \xi_j g(\mathbf{x}^i) \cdot g(\mathbf{x}^j) = \sum_{i=1}^{n} \xi_i g(\mathbf{x}^i) \sum_{j=1}^{n} \xi_j g(\mathbf{x}^j) = \left(\sum_{i=1}^{n} \xi_i g(\mathbf{x}^i) \right)^2 \geq 0$

$\square$

k_1 et k_2 semi-défini positif $\Rightarrow k_1 + k_2$ semi-défini positif.

- $k(\mathbf{x}, \mathbf{y}) = k_1(\mathbf{x}, \mathbf{y}) + k_2(\mathbf{x}, \mathbf{y}) = k_1(\mathbf{y}, \mathbf{x}) + k_2(\mathbf{y}, \mathbf{x}) = k(\mathbf{y}, \mathbf{x})$
- $\sum_{\substack{i=1\\j=1}}^{n} \xi_i \xi_j k(\mathbf{x}^i, \mathbf{x}^j) = \sum_{\substack{i=1\\j=1}}^{n} \xi_i \xi_j k_1(\mathbf{x}^i, \mathbf{x}^j) + \sum_{\substack{i=1\\j=1}}^{n} \xi_i \xi_j k_2(\mathbf{x}^i, \mathbf{x}^j) \geq 0$

□

k_1 et k_2 semi-défini positif $\Rightarrow k_1 \cdot k_2$ semi-défini positif.

□

k_1 semi-défini positif $\Rightarrow k = \exp(k_1)$ est semi-défini positif.

$$\sum_{\substack{i=1\\j=1}}^{n} \xi_i \xi_j k(\mathbf{x}^i, \mathbf{x}^j) = \sum_{\substack{i=1\\j=1}}^{n} \xi_i \xi_j \sum_{r=1}^{+\infty} \frac{k_1^r(\mathbf{x}^i, \mathbf{x}^j)}{r!} = \sum_{r=1}^{+\infty} \frac{1}{r!} \sum_{\substack{i=1\\j=1}}^{n} \xi_i \xi_j k_1^r(\mathbf{x}^i, \mathbf{x}^j) \geq 0$$

□

$k(\mathbf{x},\mathbf{y}) = \langle \mathbf{x},\mathbf{y}\rangle^p$ noyau polynomial d'ordre p.

- $k_1(\mathbf{x},\mathbf{y}) = \langle \mathbf{x},\mathbf{y}\rangle$ est un noyau
- $k_2(\mathbf{x},\mathbf{y}) = k_1(\mathbf{x},\mathbf{y}) \cdot k_1(\mathbf{x},\mathbf{y})$ est un noyau
- $\ldots$

$\square$

Correspond à l'espace de tous les monômes d'ordre p.

Variante : espace de tous les monômes jusqu'à l'ordre p.

$$k(\mathbf{x},\mathbf{y}) = \left(\langle \mathbf{x},\mathbf{y}\rangle + q\right)^p$$

$k(\mathbf{x},\mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|_2^2}{\gamma^2}\right)$ noyau radial.

$$\exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|_2^2}{\gamma^2}\right) = \exp\left(-\frac{\langle\mathbf{x},\mathbf{x}\rangle_{\mathbb{R}^d}}{\gamma^2}\right)\exp\left(-\frac{\langle\mathbf{y},\mathbf{y}\rangle_{\mathbb{R}^d}}{\gamma^2}\right)\exp\left(2\frac{\langle\mathbf{x},\mathbf{y}\rangle_{\mathbb{R}^d}}{\gamma^2}\right)$$

$$k_1(\mathbf{x},\mathbf{y}) = \exp\left(-\frac{\langle\mathbf{x},\mathbf{x}\rangle_{\mathbb{R}^d}}{\gamma^2}\right)\cdot\exp\left(-\frac{\langle\mathbf{y},\mathbf{y}\rangle_{\mathbb{R}^d}}{\gamma^2}\right) = g(\mathbf{x})\cdot g(\mathbf{y})$$

$$k_2(\mathbf{x},\mathbf{y}) = \exp\left(2\frac{\langle\mathbf{x},\mathbf{y}\rangle_{\mathbb{R}^d}}{\gamma^2}\right) = \exp\left(k_3(\mathbf{x},\mathbf{y})\right)$$

□

Espace de projection associé de dimension infinie.

Autres noyaux :

- $\mathbf{x}, \mathbf{y} \in \mathbb{R}^+$

$$k(\mathbf{x}, \mathbf{y}) = \min(\mathbf{x}, \mathbf{y})$$

- $\mathbf{x}, \mathbf{y} \in (\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.

$$k(\mathbf{x}, \mathbf{y}) = \mathbb{P}(\mathbf{x}, \mathbf{y}) - \mathbb{P}(\mathbf{x})\mathbb{P}(\mathbf{y})$$

(Mesure de similarité d'événements probabilistes)

Vecteur aléatoire $\mathbf{x} \in \mathbb{R}^n$, N observations $\mathbf{x}^i$.

Algorithme

1. Centrer
2. Matrice de covariance :

$$\mathbf{C_x} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}^i {\mathbf{x}^i}^T$$

3. Résoudre :

$$\begin{aligned} \lambda \mathbf{v} &= \mathbf{C_x v} \\ \|\mathbf{v}\|_{\mathbb{R}^n} &= 1 \end{aligned}$$

4. Projection sur k^{ieme} PC :

$$\mathbf{x}_{pc}^k = \langle \mathbf{v}^k, \mathbf{x} \rangle_{\mathbb{R}^n}$$

Projection :

$$\begin{array}{rcl} \Phi : \mathbb{R}^n & \rightarrow & \mathcal{H} \\ \mathbf{x} & \mapsto & \Phi(\mathbf{x}) \end{array}$$

Matrice de covariance dans $\mathcal{H}$:

$$\mathbf{C}_{\Phi(\mathbf{x})} = \frac{1}{N}\sum_{i=1}^{N}\Phi(\mathbf{x}^i)\Phi(\mathbf{x}^i)^T$$

Résoudre :

$$\lambda \mathbf{v}_\Phi = \mathbf{C}_{\Phi(\mathbf{x})}\mathbf{v}_\Phi$$

Projection :

$$\begin{aligned} \Phi : \mathbb{R}^n &\rightarrow \mathcal{H} \\ \mathbf{x} &\mapsto \Phi(\mathbf{x}) \end{aligned}$$

Matrice de covariance dans $\mathcal{H}$:

$$\mathbf{C}_{\Phi(\mathbf{x})} = \frac{1}{N}\sum_{i=1}^{N}\Phi(\mathbf{x}^i)\Phi(\mathbf{x}^i)^T$$

Résoudre :

$$\begin{aligned} \lambda \mathbf{v}_\Phi &= \mathbf{C}_{\Phi(\mathbf{x})}\mathbf{v}_\Phi \\ &= \frac{1}{N}\sum_{i=1}^{N}\Phi(\mathbf{x}^i)\Phi(\mathbf{x}^i)^T\mathbf{v}_\Phi \end{aligned}$$

Projection :

$$\begin{aligned}\Phi : \mathbb{R}^n &\rightarrow \mathcal{H} \\ \mathbf{x} &\mapsto \Phi(\mathbf{x})\end{aligned}$$

Matrice de covariance dans $\mathcal{H}$:

$$\mathbf{C}_{\Phi(\mathbf{x})} = \frac{1}{N}\sum_{i=1}^{N}\Phi(\mathbf{x}^i)\Phi(\mathbf{x}^i)^T$$

Résoudre :

$$\begin{aligned}\lambda\mathbf{v}_\Phi &= \mathbf{C}_{\Phi(\mathbf{x})}\mathbf{v}_\Phi \\ &= \frac{1}{N}\sum_{i=1}^{N}\Phi(\mathbf{x}^i)\Phi(\mathbf{x}^i)^T\mathbf{v}_\Phi \\ &= \frac{1}{N}\sum_{i=1}^{N}\langle\Phi(\mathbf{x}^i), \mathbf{v}_\Phi\rangle_{\mathcal{H}}\Phi(\mathbf{x}^i)\end{aligned}$$

$\mathbf{v}_\Phi$ est généré par les $\Phi(\mathbf{x}^i)$:

$$\mathbf{v}_\Phi = \sum_{i=1}^{N} \alpha_i \Phi(\mathbf{x}^i)$$

Ré-écriture :

$$\begin{aligned}
\lambda \mathbf{v}_\Phi &= \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \langle \Phi(\mathbf{x}^i), \mathbf{v}_\Phi \rangle_{\mathcal{H}} \Phi(\mathbf{x}^i) \\
\lambda \textstyle\sum_{i=1}^{N} \alpha_i \Phi(\mathbf{x}^i) &= \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \langle \Phi(\mathbf{x}^i), \textstyle\sum_{j=1}^{N} \alpha_j \Phi(\mathbf{x}^j) \rangle_{\mathcal{H}} \Phi(\mathbf{x}^i) \\
\lambda \textstyle\sum_{i=1}^{N} \alpha_i \Phi(\mathbf{x}_i) &= \tfrac{1}{N} \textstyle\sum_{\substack{i=1\\j=1}}^{N} \alpha_j \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^j) \rangle_{\mathcal{H}} \Phi(\mathbf{x}^i)
\end{aligned}$$

$\sum_{m=1}^{N} \langle \Phi(\mathbf{x}_m), . \rangle_{\mathcal{H}}$

$$\lambda \textstyle\sum_{\substack{i=1\\m=1}}^{N} \alpha_i \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^m) \rangle_{\mathcal{H}} = \tfrac{1}{N} \textstyle\sum_{\substack{i=1\\j=1\\m=1}}^{N} \alpha_j \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^j) \rangle_{\mathcal{H}} \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^m) \rangle_{\mathcal{H}}$$

$\mathbf{K}$ la matrice $N \times N$, $\mathbf{K}_{ij} := \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^j) \rangle_{\mathcal{H}}$
On obtient :

$$\lambda \mathbf{K} \boldsymbol{\alpha} = \frac{1}{N} \mathbf{K}^2 \boldsymbol{\alpha}$$

Les solutions $\boldsymbol{\alpha}$ sont trouvées par :

$$\lambda \boldsymbol{\alpha} = \frac{1}{N} \mathbf{K} \boldsymbol{\alpha}$$

Connaître $\boldsymbol{\alpha} \Rightarrow$ connaître $\mathbf{v}_{\Phi}$

Preuve.

- $\boldsymbol{\alpha} = \sum_p a_p \boldsymbol{\beta}^p$, solution de
  $\lambda \mathbf{K} \boldsymbol{\alpha} = \frac{1}{N} \mathbf{K}^2 \boldsymbol{\alpha} \iff N\lambda \sum_p a_p \mu_p \boldsymbol{\beta}^p = \sum_p a_p \mu_p^2 \boldsymbol{\beta}^p$
  $\forall p, N\lambda = \mu_p$ ou $a_p = 0$ ou $\mu_p = 0$
- $\boldsymbol{\alpha} = \sum_p a_p \boldsymbol{\beta}^p$, solution de $\lambda \boldsymbol{\alpha} = \frac{1}{N} \mathbf{K} \boldsymbol{\alpha} \iff N\lambda \sum_p a_p \boldsymbol{\beta}^p = \sum_p a_p \mu_p \boldsymbol{\beta}^p$
  $\forall p, N\lambda = \mu_p$ ou $a_p = 0$
- $\boldsymbol{\alpha} = \boldsymbol{\beta}^p | \mu_p = 0$. On a $\mathbf{K}\boldsymbol{\alpha} = 0 \Rightarrow \forall i : \sum_j \mathbf{K}_{ij} \boldsymbol{\alpha}_j = 0$
  $\Rightarrow \forall i : \sum_j \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^j) \rangle_{\mathcal{H}} \boldsymbol{\alpha}_j = 0 \Rightarrow \forall i : \langle \Phi(\mathbf{x}^i), \sum_j \boldsymbol{\alpha}_j \Phi(\mathbf{x}^j) \rangle_{\mathcal{H}} = 0$

$\square$

Le vecteur propre de $\mathbf{C}_{\Phi(\mathbf{x})}$ associé à $\boldsymbol{\beta}^p | \mu_p = 0$ n'est pas pertinent.
Projection :

$$\Phi(\mathbf{x})^k_{kpc} = \langle \mathbf{v}^k_\Phi, \Phi(\mathbf{x}) \rangle_{\mathcal{H}} = \sum_{i=1}^{N} \alpha_i^k \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}) \rangle_{\mathcal{H}}$$

Preuve.

- $\boldsymbol{\alpha} = \sum_p a_p \boldsymbol{\beta}^p$, solution de
$\lambda \mathbf{K}\boldsymbol{\alpha} = \frac{1}{N}\mathbf{K}^2\boldsymbol{\alpha} \iff N\lambda \sum_p a_p \mu_p \boldsymbol{\beta}^p = \sum_p a_p \mu_p^2 \boldsymbol{\beta}^p$
$\forall p, N\lambda = \mu_p$ ou $a_p = 0$ ou $\mu_p = 0$
- $\boldsymbol{\alpha} = \sum_p a_p \boldsymbol{\beta}^p$, solution de $\lambda \boldsymbol{\alpha} = \frac{1}{N}\mathbf{K}\boldsymbol{\alpha} \iff N\lambda \sum_p a_p \boldsymbol{\beta}^p = \sum_p a_p \mu_p \boldsymbol{\beta}^p$
$\forall p, N\lambda = \mu_p$ ou $a_p = 0$
- $\boldsymbol{\alpha} = \boldsymbol{\beta}^p | \mu_p = 0$. On a $\mathbf{K}\boldsymbol{\alpha} = 0 \Rightarrow \forall i : \sum_j \mathbf{K}_{ij}\boldsymbol{\alpha}_j = 0$
$\Rightarrow \forall i : \sum_j \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^j)\rangle_{\mathcal{H}} \boldsymbol{\alpha}_j = 0 \Rightarrow \forall i : \langle \Phi(\mathbf{x}^i), \sum_j \boldsymbol{\alpha}_j \Phi(\mathbf{x}^j)\rangle_{\mathcal{H}} = 0$

□

Le vecteur propre de $\mathbf{C}_{\Phi(\mathbf{x})}$ associé à $\boldsymbol{\beta}^p | \mu_p = 0$ n'est pas pertinent.
Projection :

$$\Phi(\mathbf{x})^k_{kpc} = \langle \mathbf{v}^k_{\Phi}, \Phi(\mathbf{x})\rangle_{\mathcal{H}} = \sum_{i=1}^{N} \alpha_i^k \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x})\rangle_{\mathcal{H}}$$

Preuve.

- $\boldsymbol{\alpha} = \sum_p a_p \boldsymbol{\beta}^p$, solution de
  $\lambda \mathbf{K}\boldsymbol{\alpha} = \frac{1}{N}\mathbf{K}^2\boldsymbol{\alpha} \iff N\lambda \sum_p a_p \mu_p \boldsymbol{\beta}^p = \sum_p a_p \mu_p^2 \boldsymbol{\beta}^p$
  $\forall p, N\lambda = \mu_p$ ou $a_p = 0$ ou $\mu_p = 0$
- $\boldsymbol{\alpha} = \sum_p a_p \boldsymbol{\beta}^p$, solution de $\lambda\boldsymbol{\alpha} = \frac{1}{N}\mathbf{K}\boldsymbol{\alpha} \iff N\lambda \sum_p a_p \boldsymbol{\beta}^p = \sum_p a_p \mu_p \boldsymbol{\beta}^p$
  $\forall p, N\lambda = \mu_p$ ou $a_p = 0$
- $\boldsymbol{\alpha} = \boldsymbol{\beta}^p | \mu_p = 0$. On a $\mathbf{K}\boldsymbol{\alpha} = 0 \Rightarrow \forall i : \sum_j \mathbf{K}_{ij}\boldsymbol{\alpha}_j = 0$
  $\Rightarrow \forall i : \sum_j \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^j)\rangle_{\mathcal{H}} \boldsymbol{\alpha}_j = 0 \Rightarrow \forall i : \langle \Phi(\mathbf{x}^i), \sum_j \boldsymbol{\alpha}_j \Phi(\mathbf{x}^j)\rangle_{\mathcal{H}} = 0$

$\square$

Le vecteur propre de $\mathbf{C}_{\Phi(\mathbf{x})}$ associé à $\boldsymbol{\beta}^p | \mu_p = 0$ n'est pas pertinent.
Projection :

$$\Phi(\mathbf{x})^k_{kpc} = \langle \mathbf{v}^k_{\Phi}, \Phi(\mathbf{x})\rangle_{\mathcal{H}} = \sum_{i=1}^{N} \alpha_i^k \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x})\rangle_{\mathcal{H}}$$

Matrice de Gram : Matrice de tout les produits scalaires possibles.

$$\mathbf{K} = \begin{pmatrix} \langle \Phi(\mathbf{x}^1), \Phi(\mathbf{x}^1) \rangle_{\mathcal{H}} & \langle \Phi(\mathbf{x}^1), \Phi(\mathbf{x}^2) \rangle_{\mathcal{H}} & \dots & \langle \Phi(\mathbf{x}^1), \Phi(\mathbf{x}^N) \rangle_{\mathcal{H}} \\ \langle \Phi(\mathbf{x}^2), \Phi(\mathbf{x}^1) \rangle_{\mathcal{H}} & \langle \Phi(\mathbf{x}^2), \Phi(\mathbf{x}^2) \rangle_{\mathcal{H}} & \dots & \langle \Phi(\mathbf{x}^2), \Phi(\mathbf{x}^N) \rangle_{\mathcal{H}} \\ \vdots & \vdots & \ddots & \vdots \\ \langle \Phi(\mathbf{x}^N), \Phi(\mathbf{x}^1) \rangle_{\mathcal{H}} & \langle \Phi(\mathbf{x}^N), \Phi(\mathbf{x}^2) \rangle_{\mathcal{H}} & \dots & \langle \Phi(\mathbf{x}^N), \Phi(\mathbf{x}^N) \rangle_{\mathcal{H}} \end{pmatrix}$$

Matrice noyau (kernel matrix) :

$$\mathbf{K} = \begin{pmatrix} k(\mathbf{x}^1, \mathbf{x}^1) & k(\mathbf{x}^1, \mathbf{x}^2) & \dots & k(\mathbf{x}^1, \mathbf{x}^N) \\ k(\mathbf{x}^2, \mathbf{x}^1) & k(\mathbf{x}^2, \mathbf{x}^2) & \dots & k(\mathbf{x}^2, \mathbf{x}^N) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}^N, \mathbf{x}^1) & k(\mathbf{x}^N, \mathbf{x}^2) & \dots & k(\mathbf{x}^N, \mathbf{x}^N) \end{pmatrix}$$

Algorithme

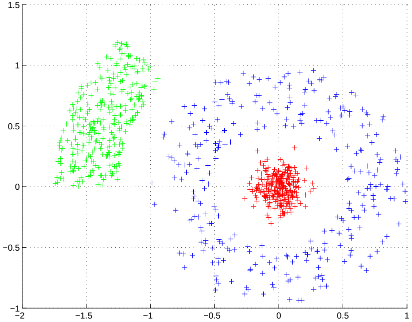
1. Construire et centrer $\mathbf{K}$
2. Diagonaliser $\mathbf{K}$
3. Normalisation des vecteurs propres :

$$\lambda_k \langle \boldsymbol{\alpha}^k, \boldsymbol{\alpha}^k \rangle = 1$$

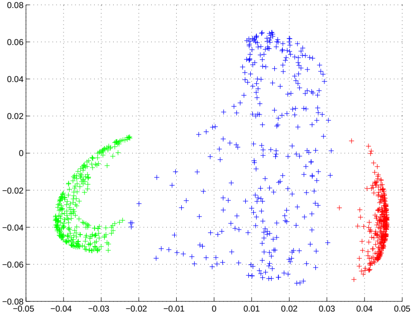
4. Projection sur k^{ieme} KPC :

$$\Phi(\mathbf{x})^k_{kpc} = \sum_{i=1}^{N} \alpha_i^k k(\mathbf{x}^i, \mathbf{x})$$

Un autre exemple...



originale



2 premières kpc

Application aux images hyperspectrales :

pc1

pc2

kpc1

kpc2

Avantages :

- Fondement théorique
- Analyse non linéaire
- Utilisation d'*a priori* pour les noyaux
- Implémentation facile

Inconvénients :

- Stockage Kernel Matrice
- Boîte noire de certain noyaux (RBF)

- Kernel ICA
- Kernel Fisher discriminant analysis
- Support Vector Machines
- Kernel Maximum Likelyhood

Méthodes à Noyaux pour L'analyse de Données

Mathieu Fauvel

gipsa-lab

19 janvier 2007