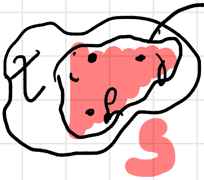
Kernel Principal Component Analysis

I. Introduction

* OBJECTIF :

Prendre en entrée une matrice K qui décrit des relations entre paires d'objets.

$K = \{K_{ij}\}_{\substack{i=1..n \\ j=1..n}}$ $\quad K_{ij}$ similarité entre i, j



$\to \Phi(S) = (aatc, atgc, tcgc)$

NON OBS.

$$K_{\text{OBS.}} = \begin{array}{c} \\ aatc \\ atgc \\ tcgc \end{array} \begin{array}{ccc} aatc & atgc & tcgc \\ \end{array}$$

$$K = \begin{array}{cccc} & aatc & atgc & tcgc \\ aatc & 1 & 0.5 & 0.5 \\ atgc & 0.5 & 1 & 0.6 \\ tcgc & 0.5 & 0.6 & 1 \end{array}$$

OBS.

• Idée : Definir un fonction de comparaison $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ pour representer les similarité dans $S \subset \mathcal{X}$

II Noyaux définis positifs

Un noyau défini positif sur un ensemble $\mathcal{X}$ est une fonction $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ symétrique :

① $\forall (x, x') \in \mathcal{X}^2, \quad K(x, x') = K(x', x)$

② $\forall n \in \mathbb{N}$

$(x_1, \dots, x_n) \in \mathcal{X}^n$ et

$(a_1, \dots, a_n) \in \mathbb{R}^n$

$$\sum_{i=1}^{n} \sum_{j=}^{n} a_i a_j K(x_i, x_j) \geq 0$$

La matrice $K = (K_{ij})_{\substack{i=1..n \\ j=1..n}}$ telle $K_{ij} = K(x_i, x_j)$, est symétrique et définie positive.

• Exemple $\quad \mathcal{X} = \mathbb{R}^d \quad K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

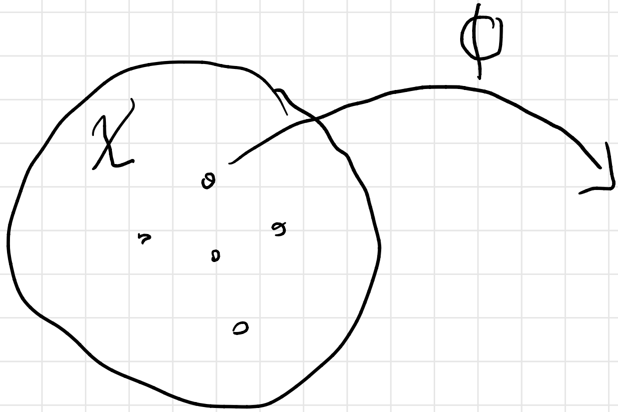
$$K(x_i, x_j) = \langle x_i, x_j \rangle_{\mathbb{R}^d} = \sum_{k=1}^{d} x_{ik} x_{jk} = x_i^T x_j$$

① $K(x_i, x_j) = K(x_j, x_i) \quad \forall i, j$

② $\forall a \in \mathbb{R}^d \quad a = \begin{pmatrix} a_1 \\ \vdots \\ a_d \end{pmatrix}$

$$\sum_{i=1}^{d} \sum_{j=1}^{d} a_i a_j \langle x_i, x_j \rangle_{\mathbb{R}^d} = \langle \sum_i a_i x_i, \sum_j a_j x_j \rangle_{\mathbb{R}^d}$$

$$= \| \sum_i a_i x_i \|^2 \geq 0$$



$K : \mathcal{X}^2 \to \mathbb{R}, \quad K(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$

① $K(x, x') = K(x', x)$

② $\forall a \in H \quad \dim(H) = n$

$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \langle \phi(x_i), \phi(x_j) \rangle_H =$

$\langle \sum_{i=1}^{n} a_i \phi(x_i), \sum_j a_j \phi(x_j) \rangle_H \geq 0$

D'un point de vue matriciel

$\forall a \in \mathbb{R}^n \quad a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ et $\phi(x_i) \in H$
$\phi(x_j) \in H$

Si K est un noyau d.p alors la matrice

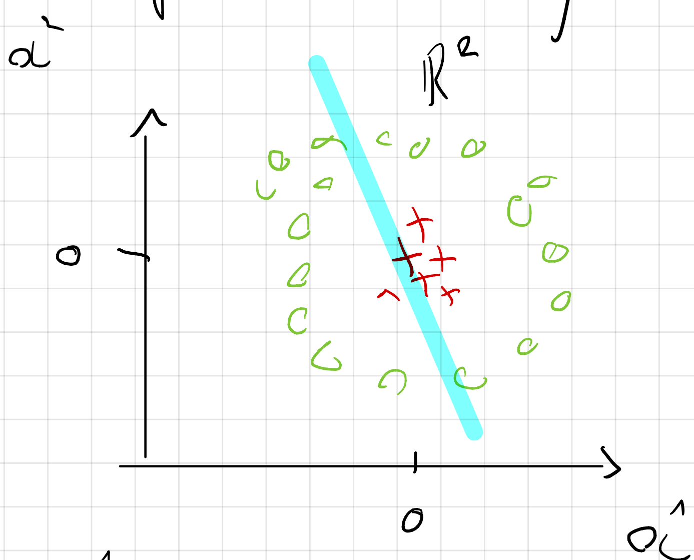
$$K = (K_{ij}) = \left(\langle \phi(x_i), \phi(x_j) \rangle_H \right)_{i=1..n}$$

vérifie $a^T K a \geq 0$.

K est donc une matrice définie positive.

III) Astuce du noyau.

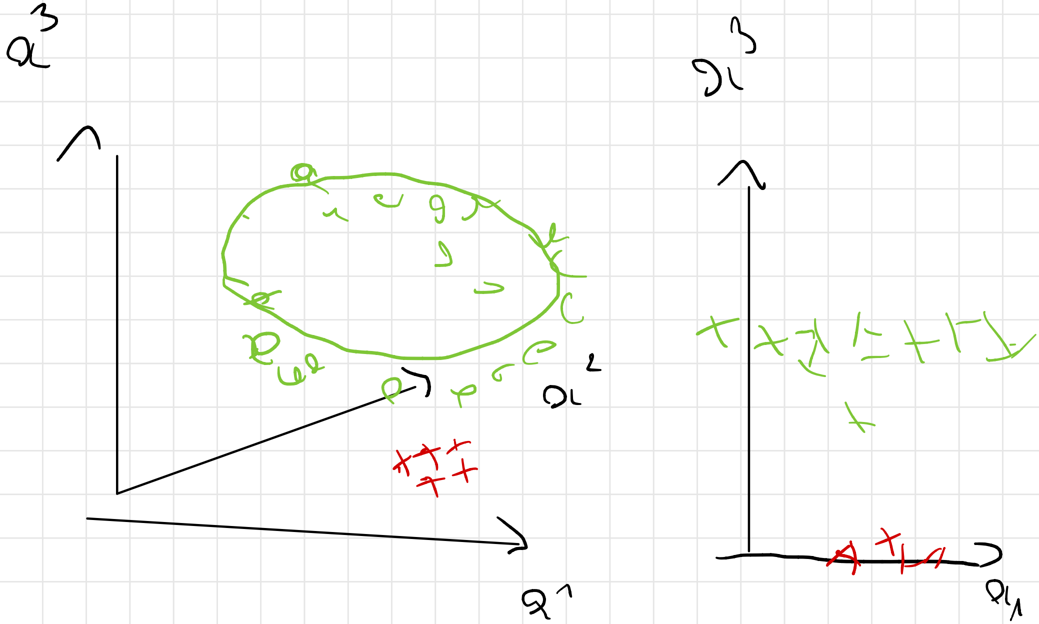
Remplacer le produit scalaire classique par un noyau quelconque.



- Apprentissage supervisé
  2 classes non linéairement séparables.

$$\phi\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} x^1 \\ x^2 \\ x^3 = (x^1)^2 + (x^2)^2 \end{pmatrix}$$

$\phi : \underset{\mathbb{R}^2}{\mathcal{X}} \longrightarrow \mathbb{R}^3$



Une méthode linéaire peu efficace dans $\mathcal{X}(\mathbb{R}^2)$, peut devenir très compétitive dans $\mathcal{H}$ (après transformation ϕ.

Astuce du noyau :

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

Exemple : $\|\phi(x_i) - \phi(x_j)\|_H^2 =$

$$< \phi(x_i) - \phi(x_j) , \phi(x_i) - \phi(x_j) >_H =$$

$$<\phi(x_i), \phi(x_i)>_H + <\phi(x_j), \phi(x_j)>_H - 2<\phi(x_i), \phi(x_j)>_H =$$

$$K_{ii} + K_{jj} - 2 K_{ij}$$

$\boxed{TV}$ KPCA (ACP Kernelisée)

$\mathcal{X}$: espace initial

$\phi : \mathcal{X} \longrightarrow \mathcal{H}$

Soit $S = (x_1, \dots, x_n) \,/\, x_i \in \mathcal{X}$

$\phi(x_i) \in \mathcal{H}$

Vecteur moyenne $\mu = \frac{1}{n} \sum_{i=1}^{n} \phi(x_i) = \underline{0}$

Matrice de covariance $S = \frac{1}{n} \sum_{i=1}^{n} \phi(x_i) \phi(x_j)^T$

Vecteurs propres $Sv = \lambda v$

Proposition

v peut être exprimé comme une combinaison linéaire

$$v = \sum_{i=1}^{n} \alpha_i \phi(x_i)$$

Preuve :

$$Sv = \underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} \phi(x_i) \overbrace{\phi(x_i)^T \right) v}^{\in \mathbb{R}}}_{S}$$

$$= \lambda \quad v$$

Remarquons que $\phi(x_i)^T v = \underbrace{\langle \phi(x_i), v \rangle_H}_{\in \mathbb{R}}$

Soit $\alpha_i = \frac{1}{n\lambda} \langle \Phi(x_i), v \rangle_{\mathcal{H}}$

$$\lambda v = \sum_{i=1}^{n} \alpha_i \Phi(x_i) \lambda$$

d'où $v = \sum_{i=1}^{n} \alpha_i \Phi(x_i)$ cqfd.

Comment trouver les α_i ?

Soit v_j le v.p associé à λ_j :

$$S v_j = \lambda_j v_j$$

$$\Phi(x_k)^T \Big(\underbrace{\frac{1}{n} \sum_{i=1}^{n} \Phi(x_i) \Phi(x_i)^T}_{S} \sum_{\ell=1}^{n} \alpha_{j\ell} \Phi(x_\ell) \Big) = \Phi(x_k)^T \lambda_j \overbrace{\sum_{i=1}^{n} \alpha_{ji} \Phi(x_i)}^{v_j}$$

$$\frac{1}{n} \sum_{i=1}^{n} K(x_k, x_i) \sum_{\ell=1}^{n} \alpha_{j\ell} K(x_i, x_\ell) = \lambda_j \sum_{i=1}^{n} \alpha_{ji} \underbrace{K(x_k, x_i)}_{\langle \Phi(x_k), \Phi(x_i) \rangle}$$

Soit $\alpha_j = \begin{pmatrix} \alpha_{j1} \\ \vdots \\ \alpha_{jn} \end{pmatrix}$ et $K(x_k, \cdot) = \begin{pmatrix} K(x_1, x_k) \\ \vdots \\ K(x_n, x_k) \end{pmatrix}$

$$\frac{1}{n} \sum_{i=1}^{n} K(x_i, x_k) K(x_i, \cdot)^T \alpha_j = \lambda_j \alpha_j^T K(x_k, \cdot)$$

Pour tous les indices $k \in \{1, \dots, n\}$

en résumé $K^2 \alpha_j = n \lambda_j K \alpha_j$

α_j vérifie $\boxed{K \alpha_j = n \lambda_j \alpha_j}$

diagonaliser K produit les vecteurs α_j qui nous donnent accès aux vecteurs v_j qui sont les vecteurs propres de S.

Pour l'ACP, on a $v_j^T v_j = 1$

$$\left\langle \sum_{k=1}^{n} \alpha_{jk} \phi(x_k)^T, \sum_{l=1}^{n} \alpha_{jl} \phi(x_l) \right\rangle = 1$$

$$\alpha_j^T K \alpha_j = 1$$

$$< \alpha_j^T \begin{pmatrix}\phi(x_1)\\ \vdots \\ \phi(x_n)\end{pmatrix}, \begin{pmatrix}\phi(x_1)\\ \vdots \\ \phi(x_n)\end{pmatrix}^T \alpha_j > =$$

$$\alpha_j^T \quad K \quad \alpha_j = 1$$

ACP dans H
$$\begin{cases} \text{Trouver } \alpha_j : & K\alpha_j = \lambda_j n \alpha_j \\ & \alpha_j^T K \alpha_j = 1 \end{cases} \Rightarrow \text{Diagonaliser } K \,!$$

- La projection de $\phi(x_k)$ sur v_j est

$$\phi(x_k)^T v_j = \sum_{i=1}^{n} \alpha_{ji} \overbrace{\phi(x_i)^T \phi(x_k)}^{K_{ik}}$$

$$= \sum_{i=1}^{n} K_{ik} \times \alpha_{jk}$$

$$\begin{pmatrix}\phi(x_1)^T\\ \vdots \\ \phi(x_n)^T\end{pmatrix} v_j = K\alpha_j \quad \forall j$$

* Nous avons supposé

$$\mu = 0 = \frac{1}{n}\sum_i \phi(x_i)$$

* Comment centrer dans $\mathcal{H}$.

on considère simplement

$$\tilde{K} = K - 2\,\mathbb{1}_{(1/n)} K + \mathbb{1}_{(1/n)} K \mathbb{1}_{(1/n)}$$

où $\mathbb{1}_{(1/n)} = \underbrace{\begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}}_{n}\Big\}\, n$

à la place de K

1. ~~Calc~~ Calcul de $\tilde{K}$ à partir de K
2. Résoudre $\tilde{K}\alpha_j = \lambda_j n\, \alpha_j \;\forall j$
3. Projection de $\phi(x_i)$ sur les v_j

$K\alpha_j$