

KECM : Une version noyau de l'algorithme évidentiel des c -moyennes

KECM : Evidential c -means in a kernel-induced feature space

M.-H. Masson¹

T. Dencœur²

¹ Université de Picardie Jules Verne

² Université de Technologie de Compiègne

Laboratoire Heudiasyc, UMR CNRS 6599

BP 20529 60205 Compiègne

mmasson@hds.utc.fr

Résumé :

Nous introduisons dans cet article un nouvel algorithme de classification automatique intitulé KEKM. Il s'agit d'une version à base de noyaux de l'algorithme ECM (algorithme évidentiel des c -moyennes). Pour formuler cet algorithme, nous montrons dans un premier temps qu'ECM peut être réécrit pour prendre pouvoir traiter des données relationnelles, ou de manière équivalente, des produits scalaires dans l'espace d'entrée. Puis, remplaçant les produits scalaires par des fonctions noyaux ("truc du noyau"), nous en déduisons l'algorithme KEKM. Grâce à une transformation non linéaire des données, KEKM est capable de détecter des classes de formes variées et de fournir, tout comme ECM, une partition évidentielle généralisant les partitions probabilistes ou floues. Notre approche est illustrée sur plusieurs jeux de données et le choix automatique de la largeur de bande du noyau Gaussien est étudié.

Mots-clés :

Classification, théorie des fonctions de croyance, méthodes à noyaux, données relationnelles.

Abstract:

A new clustering algorithm, called KEKM (kernel evidential c -means) is introduced. It is a kernelized version of ECM (Evidential c -means), a clustering algorithm based on the Dempster-Shafer theory of belief functions. We first show that ECM can be reformulated to handle relational data, or, equivalently, dot products in the original space. Then, replacing the dot products by kernel functions, KEKM is easily derived. Thanks to the non linear transformation of the data, the algorithm allows us to detect arbitrarily shaped clusters and, as ECM, generates an evidential partition, a new concept that extends the existing concepts of hard, probabilistic and fuzzy partitions. The approach is illustrated using a synthetic data set. The automatic choice of the Gaussian kernel bandwidth is also investigated.

Keywords:

Clustering, belief functions theory, kernel methods, relational clustering.

1 Introduction

La classification automatique vise à regrouper des objets similaires en classes. La similarité entre objets est soit calculée comme une distance entre des attributs décrivant les objets (données individus-variables), soit accessible directement sous forme d'une matrice de similarité ou de dissimilarité (données relationnelles). Parmi les méthodes de classification automatique les plus populaires, on trouve l'algorithme des c -moyennes flou [1] (FCM) pour les données individus-variables et sa contrepartie relationnelle RFCM [2]. Ces méthodes conduisent à une partition floue, l'appartenance de chaque objet à chaque classe étant caractérisée par une valeur réelle dans l'intervalle $[0;1]$. De nombreuses extensions au modèle initial de Bezdek ont été proposées, parmi lesquelles l'algorithme possibiliste de Krishnapuram et Keller [3] ou encore l'algorithme de Noise Clustering (NC) de Davé [4] qui permettent de détecter des objets atypiques. Plus récemment a été introduit un nouveau concept de partition, la partition crédale, qui généralise les partitions possibilistes et floues. Deux algorithmes, EVCLUS et ECM [5, 6, 7], ont été formulés pour inférer, le premier à partir de données relationnelles, le second à partir de données individus-variables, ces partitions crédales. Une autre voie d'amélioration du modèle de Bezdek a résidé dans l'utilisation de noyaux [8, 10, 9]. Cette approche revient à

appliquer une transformation non linéaire aux données dans l'espace d'entrée, permettant par là même de détecter des classes de forme arbitraire. Il est possible d'obtenir une version à base de noyaux de tout algorithme qui peut se formuler en termes de produits scalaires dans l'espace initial. Nous montrons dans cet article comment reformuler ECM pour travailler avec des produits scalaires (fournissant donc une version relationnelle de ECM appelée RECM), puis comment, en remplaçant les produits scalaires par des fonctions noyaux, obtenir la version KECM de l'algorithme original.

2 Rappels sur ECM

2.1 Partition crédale

La notion de partition crédale repose sur le cadre théorique des fonctions de croyance [11] et sur l'interprétation qu'en fait Smets dans le modèle des croyances transférables [12]. On considère un ensemble de c classes $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ et n objets à classer. Nous proposons de représenter la connaissance partielle que nous avons sur l'appartenance d'un objet à une des classes par une fonction de masse m de 2^Ω dans $[0, 1]$, vérifiant $\sum_{A \subseteq \Omega} m(A) = 1$. Un sous-ensemble A recevant une masse non nulle est appelé un élément focal de m . L'ignorance totale est représentée par une fonction de masse vide ($m(\Omega) = 1$). La quantité $m(\emptyset)$ est interprétée comme la part de croyance dans le fait que la classe de l'objet n'appartient pas à Ω . Cette représentation de l'information permet de rendre compte d'une grande variété de situations comme l'illustre l'exemple suivant.

EXEMPLE 1 Considérons un ensemble de $n = 5$ individus et $c = 3$ classes. On suppose que les masses décrivant la répartition des individus dans les classes sont données dans le tableau 1. La classe de l'individu 2 est connue avec certitude, alors que l'incertitude pour l'individu 4 est totale ; le cas de l'individu 3 correspond à une situation de connaissance partielle (notons que m_3 est une masse Bayésienne car

les éléments focaux sont des singletons) ; enfin, la masse $m_1(\emptyset) = 1$ indique que la classe l'individu 1 n'appartient pas à Ω (l'individu 1 est un individu atypique).

Tableau 1 – Exemple de partition crédale

A	$m_1(A)$	$m_2(A)$	$m_3(A)$	$m_4(A)$
\emptyset	1	0	0	0
$\{\omega_1\}$	0	0	0.2	0
$\{\omega_2\}$	0	1	0.4	0
$\{\omega_1, \omega_2\}$	0	0	0	0
$\{\omega_3\}$	0	0	0.4	0
$\{\omega_1, \omega_3\}$	0	0	0	0
$\{\omega_2, \omega_3\}$	0	0	0	0
Ω	0	0	0	1

On définit une *partition crédale* comme un n -uplet $M = (m_1, \dots, m_n)$ où chaque m_i est une fonction de masse définie sur Ω . On retrouve comme cas particuliers de cette partition crédale les partitions floues, possibilistes ou encore celle de Davé.

2.2 L'algorithme ECM

ECM détermine, pour chaque objet i , les quantités $m_{ij} = m_i(A_j)$ ($A_j \neq \emptyset, A_j \subseteq \Omega$) de telle sorte que m_{ij} soit faible (resp. élevée) si la distance d_{ij} entre i et l'élément focal A_j est grande (resp. faible). La distance entre un objet et tout sous ensemble non vide de Ω est définie comme suit : tout comme dans FCM, chaque classe ω_k est représentée par un centre $\mathbf{v}_k \in \mathbb{R}^p$; puis, à chaque sous-ensemble A_j de Ω , on associe l'isobarycentre $\bar{\mathbf{v}}_j$ des représentants des classes formant A_j . La distance entre l'objet i et A_j est définie comme la distance Euclidienne entre i et $\bar{\mathbf{v}}_j$.

Pour déterminer la partition crédale $M = (m_1, \dots, m_n)$ et la matrice V de taille $(c \times p)$ des centres des classes, ECM minimise le critère suivant :

$$\sum_{i=1}^n \sum_{\{j/A_j \neq \emptyset, A_j \subseteq \Omega\}} c_j^\alpha m_{ij}^\beta d_{ij}^2 + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta, \quad (1)$$

sous les contraintes

$$\sum_{\{j/A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij} + m_{i\emptyset} = 1 \quad \forall i = 1, n, \quad (2)$$

où $m_{i\emptyset} = m_i(\emptyset)$. Le critère utilisé est similaire à celui de Davé. L'ensemble vide est assimilé à une classe de bruit supposée à une distance fixe δ de chaque objet. Le paramètre δ est utilisé pour contrôler le taux d'objets considéré comme atypiques. Le paramètre β sert à contrôler la dureté de la partition. Les coefficients (c_j^α) permettent la pénalisation des sous-ensembles de Ω de forte cardinalité, l'exposant α réglant le degré de pénalisation. Une procédure de minimisation alternée du critère est mise en oeuvre. On considère dans un premier temps que V est fixe. Les conditions d'optimalité pour M sont les suivantes (voir pour plus de détails les références [6] et [7]) :

$$m_{ij} = \frac{c_j^{-\alpha/(\beta-1)} d_{ij}^{-2/(\beta-1)}}{\sum_{A_k \neq \emptyset} c_k^{-\alpha/(\beta-1)} d_{ik}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}} \quad (3)$$

$$\forall i = 1, n \quad \forall j/A_j \subseteq \Omega, A_j \neq \emptyset \quad (4)$$

et

$$m_{i\emptyset} = 1 - \sum_{A_j \neq \emptyset} m_{ij} \quad \forall i = 1, n. \quad (5)$$

Dans un deuxième temps, on considère que M est fixe. La détermination de V est un problème d'optimisation non contraint. Soit $X = (x_{ij})$ la matrice d'entrée de taille $(n \times p)$ telle que que x_{ij} représente le j^{eme} attribut de l'objet i . En annulant les dérivées partielles du critère par rapport aux v_k , on est amené à résoudre pour V le système linéaire suivant :

$$HV = UX, \quad (6)$$

où H et U sont deux matrices de tailles respectives $(c \times c)$ et $(c \times n)$, définies par :

$$H_{lk} = \sum_i \sum_{A_j \neq \emptyset} c_j^{\alpha-2} m_{ij}^\beta s_{lj} s_{kj} \quad k, l = 1, c, \quad (7)$$

et

$$U_{li} = \sum_{A_j \neq \emptyset} c_j^{\alpha-1} m_{ij}^\beta s_{lj} \quad l = 1, c \quad i = 1, n. \quad (8)$$

avec ,

$$s_{kj} = \begin{cases} 1 & \text{if } \omega_k \in A_j \\ 0 & \text{sinon.} \end{cases} \quad (9)$$

Une partition crédale est une représentation riche qui porte beaucoup d'information sur les données. Parmi plusieurs outils d'interprétation proposés dans [7], une façon directe de synthétiser cette information est de calculer la *partition crédale nette*. Pour cela, on affecte chaque objet au sous-ensemble de Ω qui reçoit le plus de masse. D'autres outils sont proposés dans [7], comme la conversion en partition floue ou la détermination d'estimations inférieure et supérieure des classes. Signalons enfin la possibilité de combiner plusieurs partitions issues de différentes sources.

3 Formulation relationnelle d'ECM

3.1 Principe

On suppose dans cette partie que les données d'entrée consistent en une matrice $\Delta = (\delta_{ii'})$ de dissimilarités entre les objets ($\delta_{ii'}$ est supposé être égal à la distance euclidienne entre les objets i et j bien que la valeur des attributs des objets soit inconnue). Pour formuler une version relationnelle d'ECM, on doit exprimer les équations de mises à jour des masses seulement en termes de distances ou, de manière équivalente, en termes de produits scalaires entre les \mathbf{x}_i , puisque, en supposant sans perte de généralité la configuration des objets centrée sur l'origine, la relation suivante existe :

$$\mathbf{x}_i^t \mathbf{x}_{i'} = \frac{1}{2}(\delta_i + \delta_{i'} - \delta_{ii'}), \quad (10)$$

avec δ_i and $\delta_{..}$ définis, respectivement, comme :

$$\delta_i = \frac{1}{n} \sum_j \delta_{ij}, \quad (11)$$

et

$$\delta_{..} = \frac{1}{n^2} \sum_j \sum_l \delta_{jl}. \quad (12)$$

Notons que les équations (3) nécessitent le calcul de la distance Euclidienne d_{ij} entre chaque objet \mathbf{x}_i et le barycentre associé à chaque ensemble non vide A_j de Ω . Cette distance peut

facilement se calculer en termes de produits scalaires dans l'espace d'entrée, en effet :

$$\begin{aligned}
d_{ij}^2 &= (\mathbf{x}_i - \frac{1}{c_j} \sum_{k=1}^c s_{kj} \mathbf{v}_k)^t (\mathbf{x}_i - \frac{1}{c_j} \sum_{l=1}^c s_{lj} \mathbf{v}_l) \\
&= \mathbf{x}_i^t \mathbf{x}_i - \frac{2}{c_j} \sum_{k=1}^c s_{kj} \mathbf{x}_i^t \mathbf{v}_k \\
&\quad + \frac{1}{c_j^2} \sum_{k=1}^c \sum_{l=1}^c s_{kj} s_{lj} \mathbf{v}_k^t \mathbf{v}_l. \tag{13}
\end{aligned}$$

Seuls les produits scalaires $\mathbf{x}_i^t \mathbf{x}_i$ sont directement accessibles grâce à l'équation (10). Il faut donc pouvoir calculer les autres produits $\mathbf{x}_i^t \mathbf{v}_k$ et $\mathbf{v}_k^t \mathbf{v}_l$. Il se trouve qu'on peut les déduire de l'équation (6). Pour le montrer, on utilise dans ce qui suit les notations suivantes : soient $W = (w_{ii'})$ la matrice de taille $(n \times n)$ des produits scalaires des \mathbf{x}_i ($w_{ii'} = \mathbf{x}_i^t \mathbf{x}_{i'}$), $Q = (q_{kk'})$ la matrice de taille $(c \times c)$ des produits scalaires entre les \mathbf{v}_k ($q_{kk'} = \mathbf{v}_k^t \mathbf{v}_{k'}$) et $R = (r_{ki})$ la matrice de taille $(c \times n)$ des produits scalaires entre les \mathbf{v}_k et les \mathbf{x}_i ($r_{ki} = \mathbf{v}_k^t \mathbf{x}_i$). Notons que $W = XX^t$, $R = VX^t$ et $Q = VV^t$. En partant de (6), les matrices Q et R peuvent être déterminées en deux étapes :

1. **Détermination de R.** En multipliant à droite chaque côté de l'égalité (6) par X^t , on obtient :

$$HR = UW, \tag{14}$$

de telle sorte que R est obtenu par la résolution du système linéaire (14).

2. **Détermination de Q.** En multipliant à droite chaque côté de l'égalité (6) par V^t , on voit que Q est à son tour solution du système linéaire suivant :

$$HQ = UR^t. \tag{15}$$

Les matrices R et Q étant calculées, on en déduit les d_{ij} et donc les masses m_{ij} . L'algorithme RECM démarre avec une initialisation aléatoire de M , puis itère jusqu'à ce que la norme de la différence entre les matrices M obtenues lors de deux itérations successives soit inférieure à un seuil ϵ .

3.2 Exemple

Nous illustrons l'utilisation de RECM à l'aide d'une base de données réelles, la base USPS qui porte sur la reconnaissance de l'écriture manuscrite de chiffres (<http://www.kernel-machines.org/data.html>). Cette base contient 9298 chiffres de 0 à 9. Chaque chiffre est donné sous forme d'une imagerie en niveau de gris de taille 16×16 . A des fins illustratives, nous avons sélectionné au hasard 25 représentants des classes "0", "1" et "7". Ces trois classes ont été retenues car elles possèdent une structure assez claire dans l'espace, comme l'illustre la représentation par positionnement multidimensionnel (MDS) donnée en figure 1. En suivant l'approche proposée dans [13], un indice de dissimilarité $\delta_{ij} = 1 - s_{ij}$ a été calculé entre chaque paire d'images (binarisées) en utilisant le score de Simpson s_{ij} défini par :

$$s_{ij} = \frac{a}{\min(a + b, a + c)}, \tag{16}$$

où a compte le nombre de pixels égaux à 1 dans les deux images, b le nombre de pixels égaux à 0 dans i et 1 dans l'image j , et c le nombre de pixels égaux à 1 dans l'image i et à 0 dans l'image j . Les résultats de RECM, obtenus avec les valeurs expérimentales $c=3$, $\alpha = 0.5$, $\beta = 1.5$, $\delta^2 = 0.25$, $\epsilon = 10^{-3}$, sont présentés sur les figures 2 à 5. On peut voir que les trois classes naturelles sont bien reconstituées. De plus, l'affectation de certains chiffres à des classes non singletons permet de détecter des représentants ambigus ou atypiques.

4 Une version noyau de ECM

4.1 Principe

Les approches à base de noyaux sont devenues populaires depuis plusieurs années pour résoudre des problèmes d'apprentissage supervisé ou non. Plusieurs méthodes ont ainsi vu le jour, comme les séparateurs à vastes marges (SVM) [14, 15], l'analyse en composantes principales à noyau [16] ou encore l'analyse discriminante de Fisher à noyau [17]. Toutes

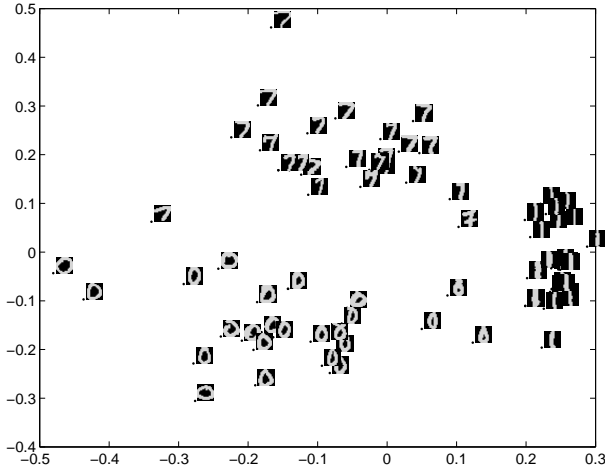


Figure 1 – Représentation par MDS du jeu de données USPS.



Figure 2 – Chiffres affectés à $\{\omega_1\}$ dans la partition crédale nette.

ces méthodes se fondent sur la possibilité de convertir une méthode linéaire en une méthode non linéaire en utilisant une transformation implicite ϕ de l'espace des données \mathcal{X} vers un espace de dimension plus élevée \mathcal{F} . Le théorème de Mercer pose que, sous certaines conditions imposées à un noyau $K(\mathbf{x}, \mathbf{y})$ dans \mathcal{X} , il existe une fonction ϕ de \mathcal{X} dans un espace \mathcal{F} de taille possiblement plus élevée telle que :

$$K(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^t \phi(\mathbf{y}). \quad (17)$$

Un noyau qui remplit les conditions est le noyau gaussien :

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2}\right), \quad (18)$$

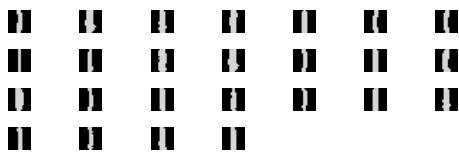


Figure 3 – Chiffres affectés à $\{\omega_2\}$ dans la partition crédale nette.



Figure 4 – Chiffres affectés à $\{\omega_3\}$ dans la partition crédale nette.

$\{\omega_1, \omega_2\}$		—	—	—
$\{\omega_2, \omega_3\}$			—	—
$\{\omega_1, \omega_3\}$	—	—	—	—
Ω	—	—	—	—
Empty set				

Figure 5 – Chiffres affectés à des sous-ensembles non singletons dans la partition crédale nette.

où σ est représenté la largeur de bande du noyau. L'idée centrale est que l'on peut calculer le produit scalaire dans \mathcal{F} sans explicitement connaître la transformation ϕ . Toute méthode linéaire peut donc être transformée en une méthode non linéaire, pourvu que l'on puisse l'exprimer uniquement en termes de produits scalaires. Chaque fois qu'un produit scalaire est utilisé dans la méthode originelle, il est remplacé par la fonction noyau. Une telle approche a déjà été utilisée dans un contexte non supervisé [8, 10, 9]. Les résultats obtenus montrent que des classes de forme variée peuvent être détectées. Nous présentons dans ce qui suit une version à base de noyaux pour l'algorithme ECM.

4.2 KECM

On considère une transformation ϕ de $\mathcal{X} \subseteq \mathbb{R}^p$ vers $\mathcal{F} \subseteq \mathbb{R}^q$. Soit \mathbf{v}_k^ϕ le centre de la classe ω_k dans l'espace induit par ϕ et V^ϕ la matrice formée en ligne par les centres \mathbf{v}_k^ϕ . Pour construire une version noyau d'ECM, on modifie le critère de la manière suivante. On cherche à minimiser :

$$\sum_{i=1}^n \sum_{\{j/A_j \neq \emptyset, A_j \subseteq \Omega\}} c_j^\alpha m_{ij}^\beta d_{ij}^{\phi^2} + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta, \quad (19)$$

sous les contraintes

$$\sum_{\{j/A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij} + m_{i\emptyset} = 1 \quad \forall i = 1, n, \quad (20)$$

où $d_{ij}^{\phi^2}$ est la distance entre l'objet i et l'élément focal $A_j \subseteq \Omega$ dans l'espace induit par ϕ :

$$\begin{aligned} d_{ij}^{\phi^2} &\triangleq \left\| \phi(\mathbf{x}_i) - \frac{1}{c_j} \sum_{k=1}^c s_{kj} \mathbf{v}_k^\phi \right\|^2 \\ &= \phi(\mathbf{x}_i)^t \phi(\mathbf{x}_i) - \frac{2}{c_j} \sum_{k=1}^c s_{kj} \phi(\mathbf{x}_i)^t \mathbf{v}_k^\phi \\ &\quad + \frac{1}{c_j^2} \sum_{k=1}^c \sum_{l=1}^c s_{kj} s_{lj} \mathbf{v}_k^{\phi t} \mathbf{v}_l^\phi. \end{aligned} \quad (21)$$

Les paramètres dans (19) ont la même signification qu'auparavant, sauf δ , qui représente maintenant la distance des objets à la classe de bruit dans l'espace transformé.

La minimisation du critère est réalisée grâce à un processus d'optimisation alternée similaire à celui utilisé pour ECM. Dans un premier temps, on considère que V^ϕ est fixe. Les équations de mise à jour pour M sont facilement obtenues en dérivant le Lagrangien par rapport à M . Ce sont les mêmes équations que pour ECM, en remplaçant d_{ij}^2 par $d_{ij}^{\phi^2}$ dans (3). Dans un second temps, M est fixé. En annulant les dérivées partielles du critère par rapport aux centres \mathbf{v}_k^ϕ , on obtient les équations suivantes :

$$\begin{aligned} \sum_i \phi(\mathbf{x}_i) \sum_{A_j \neq \emptyset} c_j^{\alpha-1} m_{ij}^\beta s_{lj} &= \\ \sum_k \mathbf{v}_k^\phi \sum_i \sum_{A_j \neq \emptyset} c_j^{\alpha-2} m_{ij}^\beta s_{lj} s_{kj} & \quad l = 1, c. \end{aligned} \quad (22)$$

Ces équations peuvent se mettre sous la forme :

$$HV^\phi = U\Phi, \quad (23)$$

où Φ désigne la matrice de taille $(n \times q)$ dont les lignes sont les images $\phi(\mathbf{x}_i)$ des objets dans l'espace induit. Notons que (23) est similaire à (6). On suit ensuite l'approche développée au paragraphe précédent : soit $W^\phi = (w_{ii'}^\phi)$

et $R^\phi = (r_{ki}^\phi)$ les matrices des produits scalaires dans l'espace transformé (de taille $(n \times n)$ et $(c \times n)$, respectivement) telles que $w_{ii'}^\phi = \phi(\mathbf{x}_i)^t \phi(\mathbf{x}_{i'})$ et $r_{ki}^\phi = \mathbf{v}_k^{\phi t} \phi(\mathbf{x}_i)$; soit $Q^\phi = (q_{kk'}^\phi)$ la matrice de taille $(c \times c)$ des produits scalaires \mathbf{v}_k^ϕ ($q_{kk'}^\phi = \mathbf{v}_k^{\phi t} \mathbf{v}_{k'}^\phi$). En utilisant ces notations, on a les relations suivantes : $W^\phi = \Phi\Phi^t$, $R^\phi = V^\phi\Phi^t$ et $Q^\phi = V^\phi V^{\phi t}$. En multipliant à droite (23) par Φ^t , on obtient :

$$HR^\phi = UW^\phi. \quad (24)$$

Appliquant l'astuce du noyau, nous remplaçons $\phi(\mathbf{x}_i)^t \phi(\mathbf{x}_{i'})$ par une fonction noyau $K(\mathbf{x}_i, \mathbf{x}_{i'})$ calculée dans l'espace original, et nous trouvons R^ϕ en résolvant (24). Puis, en multipliant à droite (23) par $V^{\phi t}$, on obtient :

$$HQ^\phi = UR^{\phi t}, \quad (25)$$

qui peut être résolue pour trouver Q^ϕ .

4.3 Exemple

On considère un jeu de données synthétique utilisé par Girolami [8] pour démontrer l'efficacité des méthodes à noyaux. Le jeu de données est composé de 1129 points en dimension 2. Il y a deux classes, chacune centrée sur l'origine. L'une est distribuée de manière gaussienne, l'autre de manière uniforme autour d'un anneau de diamètre unité. Pour tester la robustesse de KECM par rapport aux points aberrants, nous avons ajouté 100 points échantillonnés suivant une loi de Student à 5 degrés de liberté centrée en $[0; 0]^t$. Une partition crédale en deux classes a été recherchée. Dans toutes les expériences, nous avons fixé β à 2, ce qui est un choix usuel pour les méthodes floues. Nous avons utilisé un noyau gaussien. La largeur de bande du noyau a été fixée empiriquement à 2. Nous rapportons ici le résultat de quelques expériences lors desquelles nous avons fait varier δ et α . Les figures 6 et 7 montrent les partitions crédales nettes obtenues avec $\alpha = 1$ et $\delta^2 = 10$ et $\alpha = 1$ et $\delta^2 = 0.8$. Dans chacun des cas, l'élément focal Ω est plus pénalisé que les singletons dans la

fonction objective, de telle sorte qu'aucun point ne lui est affecté. Lorsque δ est faible, une part importante de masse est allouée à l'ensemble vide, ce qui permet de détecter des outliers. On voit que dans les deux cas, les deux classes, non linéairement séparables, ont été correctement détectées. La figure 8 montre les résultats obtenus lorsque on fixe α à 0. Dans ce cas, tous les éléments focaux, quelle que soit leur cardinalité, sont pénalisés de la même manière. On observe que certains points, situés à la périphérie de la classe centrale, sont maintenant affectés à Ω .

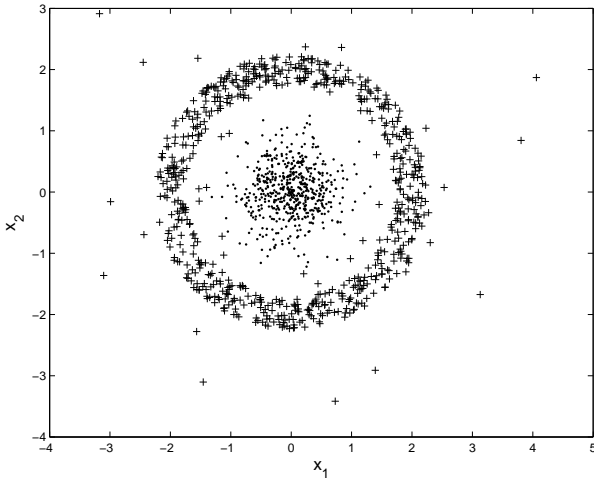


Figure 6 – Données anneau ; partition crédale nette avec $\delta^2 = 10$ et $\alpha = 1$.

Un des problèmes centraux des méthodes à base de noyaux est la sélection du paramètre qui définit le noyau, comme la largeur de bande dans le cas gaussien. Dans [6, 7], nous avons proposé un critère de validité pour choisir le nombre de classes. Il s'avère que ce critère peut également guider le choix de σ , comme nous l'illustrons ci-dessous. La définition de ce critère repose sur l'idée intuitive suivante : si le choix de c et de σ est correct, les centres vont se placer sur des zones de forte densité et la majeure partie de la masse sera allouée à des singletons de Ω . Au contraire, si ces valeurs sont mal choisies, on retrouvera une masse importante sur des éléments de cardinalité élevée ou sur l'ensemble vide. Ces éléments amènent à utiliser une mesure de non-spécificité, fondée

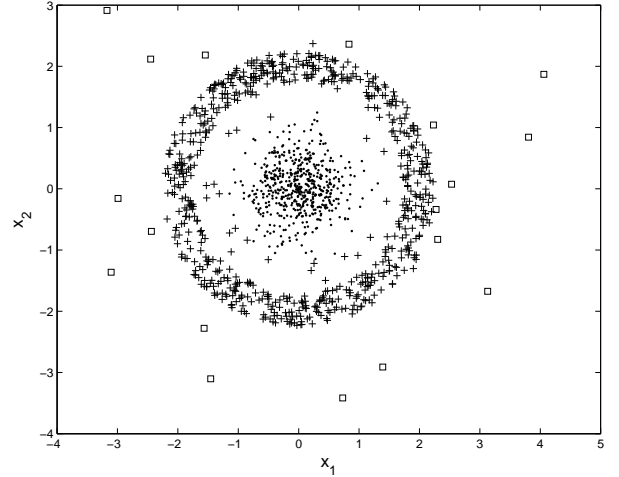


Figure 7 – Données anneau ; partition crédale nette avec $\delta^2 = 0.8$ et $\alpha = 1$; les points affectés à l'ensemble vide sont représentés par des carrés.

sur celle de Klir [18], définie par :

$$N^* \triangleq \frac{1}{n \log_2(c)} \sum_{i=1}^n \left[\sum_{A \in 2^\Omega \setminus \emptyset} m_i(A) \log_2 |A| + m_i(\emptyset) \log_2(c) \right]. \quad (26)$$

Notons que $0 \leq N^* \leq 1$. Cette mesure tend à être faible lorsque la masse est allouée à peu d'éléments focaux de faible cardinalité. Cet indice doit être minimisé. Son application est illustrée sur le jeu de données anneau. KECM a été appliqué avec $c = 2$, $\alpha = 1$, $\beta = 2$, et $\delta^2 = 10$ pour plusieurs valeurs σ . Les résultats sont présentés sur la figure 9. Il y a un minimum clair pour $\sigma = 2$, valeur retenue lors des expériences présentées précédemment.

5 Conclusion

Nous avons proposé dans cet article une nouvelle méthode de classification automatique, KECM, reposant sur le cadre théorique des fonctions de croyance. Il s'agit d'une version à base de noyaux de l'algorithme ECM permettant d'extraire une partition crédale à partir de données individu-variables. Grâce à l'introduction d'un noyau, des classes de forme

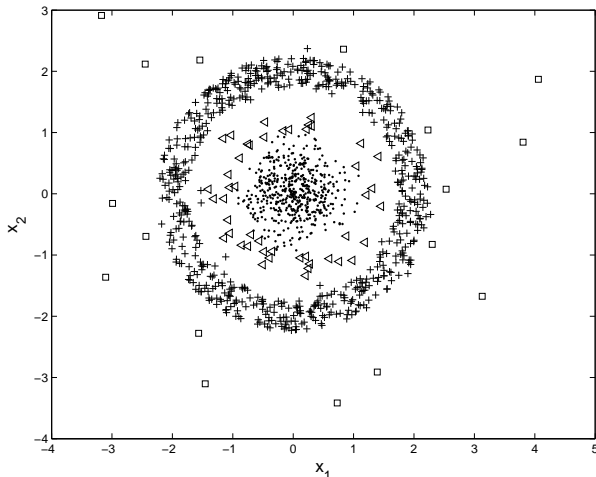


Figure 8 – Données anneau ; partition crédale nette avec $\delta^2 = 0.8$ et $\alpha = 0$; les points affectés à Ω sont représentés par des triangles.

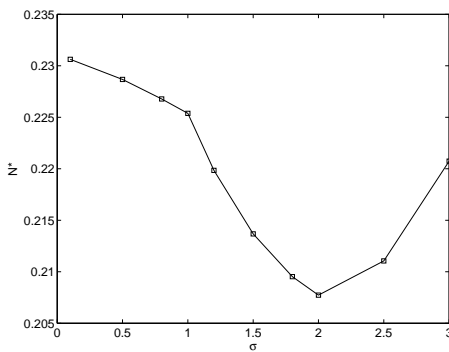


Figure 9 – Données anneau ; indice de non spécificité de la partition en fonction de la largeur de bande du noyau.

arbitraire peuvent être déterminées, comme l'illustre l'exemple de l'anneau. La grande richesse et la grande flexibilité de la partition crédale permet une analyse approfondie des données et une certaine robustesse par rapport à des données atypiques. De plus, comme étape intermédiaire, nous avons proposé un algorithme intitulé RECM dédié à l'analyse de données relationnelles.

Références

[1] J. C. Bezdek. *Pattern Recognition with fuzzy objective function algorithms.* Plenum Press, New-York, 1981.

[2] R.J. Hathaway, J.W. Davenport, and J.C. Bezdek. Relational duals of the c-means clustering algo-

rithms. *Pattern Recognition*, 22, 205-212, 1989.

[3] R. Krishnapuram and J. Keller. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2), 98-110, 1993.

[4] R.N. Davé. Clustering relational data containing noise and outliers. *Pattern Recognition Letters*, 12, 657-664, 1991.

[5] T. Denœux and M.-H. Masson. EVCLUS : Evidential CLUSTERing of proximity data. *IEEE Transactions on Systems, Man and Cybernetics Part B*, 34(1), 95-109, 2004.

[6] M.-H. Masson and T. Denœux. ECM : Algorithme évidentiel des c-moyennes. *Recontres Francophones sur la Logique Floue et ses Applications (LFA '07)*, 17-24, Nîmes, France, Novembre, 2007.

[7] M.-H. Masson and T. Denœux. ECM : An evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, doi :10.1016/j.patcog.2007.08.014, 2007.

[8] M. Girolami. Mercer kernel-based clustering in feature space. *IEEE Trans. on Neural Networks*, 13(3), 780-784, 2002.

[9] X.-H. Wu. A possibilistic c-means clustering algorithm based on kernel method. *Proceedings of the 2006 International Conf. on Communications, Circuits and Systems*, 3, 2062-2066, 2006.

[10] D.Q. Zhang and S.C. Chen. A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. *Artificial Intelligence in Medicine*, 32, 37-50, 2004.

[11] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, Princeton, N.J., 1976.

[12] P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66, 191-243, 1994.

[13] J. Laub, K.-R. Müller. Feature discovery in non-metric pairwise data. *Journal of Machine Learning Research*, 5, 801-818, 2004.

[14] V. Vapnik. *Statistical Learning Theory.* Wiley, New York, 1998.

[15] B. Schölkopf and A. Smola. *Learning with Kernels.* MIT Press, Cambridge, MA, 2002.

[16] B. Schölkopf, A. J. Smola, and K.-R. Müller. Non-linear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5), 1299-1319, 1998.

[17] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller. Fisher Discriminant Analysis with Kernels. Y.-H. Hu, J. Larsen, E. Wilson and S. Douglas, Eds. *Neural Networks for Signal Processing IX*. 41-48, IEEE, 1999.

[18] G.J. Klir, and M.J. Wierman. *Uncertainty-based Information. Elements of Generalized Information Theory.* Springer-Verlag, New-York, 1998.