Revista de Matemática: Teoría y Aplicaciones 2009 **16**(2): 241–254
cimpa – ucr issn: 1409-2433

PROPIEDADES DE UN ENFOQUE DIFUSO PARA PARTICIONAMIENTO CON DATOS BIMODALES

William Castillo* Javier Trejos†

Recibido/Received: 21 Jan 2009 — Aceptado/Accepted: 12 May 2009

Resumen

Se presenta un nuevo criterio de particionamiento difuso con datos bimodales y se formula un algoritmo para su optimización. Se estudian las propiedades de convergencia del método. El método fue probado con datos reales. De los resultados se dedujo que algunas veces con este nuevo método se pueden obtener mejores particiones bimodales que las encontradas por otros métodos.

Palabras clave: partición bimodal, algoritmo, clase difusa, minimizar, varianza explicada.

Abstract

It is presented a new fuzzy partitioning criterion for two-mode data and an algorithm is formulated for its optmization. We study the convergence properties of the method. The method was tested on real data. It is deduced that better results can be obtained with this new method than with other methods.

Keywords: Two-mode partition, algorithm, fuzzy cluster, minimization, variance accounted for.

Mathematics Subject Classification: 91C20, 03E72.

1 Introducción

En clasificación no jerárquica unimodal se busca usualmente una partición de un conjunto finito, que minimiza un criterio de clasificación. En el caso cuantitativo, este criterio es

*CIMPA, Escuela de Matemática, Universidad de Costa Rica, 2060 San José, Costa Rica. E-Mail: wcastill@cariari.ucr.ac.cr

†Misma dirección que W. Castillo. E-Mail: javier.trejos@ucr.ac.cr.

muchas veces del tipo mínimo cuadrático. Es decir,

$$L(P,V) = \sum_{k=1}^{K} \sum_{i \in I} p_{ik} \left\| x_i - v_k \right\|^2, \tag{1}$$

donde:

- $P = (p_{ik})_{p \times K}$ es la matriz cuya columna k ésima es la función indicadora de la k ésima clase de la partición P del conjunto I, con K clases.
- Para cada $i \in I$, $x_i \in I\!R^h$ es la i ésima fila de la matriz de datos X, $p \times h$, de individuos por variables continuas.
- Para cada $k \in \{1, \ldots, K\}$, $v_k \in I\!R^h$ es un vector representativo de la k ésima clase de P, por ejemplo su centro de gravedad. V es la matriz $K \times h$ cuyas filas son los K vectores v_k.

Los métodos usuales de particionamiento que minimizan el criterio $L(P,V)$ son los llamados k-medias, nubes dinámicas e ISODATA. Estos métodos están basados en búsqueda local de particiones que mejoran iterativamente el criterio hasta estabilizarse en un mínimo local.
En la sección siguiente se introduce el método de clasificación difusa unimodal con el criterio propuesto por Bezdek, J.C. ([2], [3] y [4]) el cual es una generalización del criterio (1). En la sección 3 se propone un método de clasificación difusa con datos bimodales.

2 Clasificación unimodal difusa no jerárquica

Cuando se usan clases difusas se modifica el criterio de clasificación, y el método de minimización requiere, igual que en el caso clásico, la definición de un algoritmo que converja. El siguiente criterio de clasificación difusa fue propuesto por ([2], [3] y [4]):

$$L(P,V,s) = \sum_{k=1}^{K} \sum_{i \in I} p_{ik}^s \left\| x_i - v_k \right\|^2, \tag{2}$$

donde $s \geq 1$ y la matriz P satisface las siguientes propiedades:

1. Para todo i, $\sum_{k=1}^{K} p_{ik} = 1$.
2. Para todo i, k; $p_{ik} \in [0,1]$.
3. Para todo k, $\sum_{i=1}^{n} p_{ik} > 0$.

Para un objeto i, los K números p_{ik} representan el grado de pertenencia de i a la clase k. La matriz P se llama **partición difusa**, con K clases difusas. Cada columna de P representa una clase difusa. Si $\alpha \in]0,1[$ es un umbral, entonces $A_k = \{i \in I \,|p_{ik} \geq \alpha\}$ es la clase difusa que, en adelante, se la identifica con la columna k-ésima de P.

Puede notarse que el criterio (2) es una generalización del criterio (1), puesto que las funciones indicadoras satisfacen las condiciones 1., 2. y 3.

Valores grandes del exponente s ($s = 2$ ó 3) tienen el efecto de agrandar la diferencia (con respecto, por ejemplo, a $s = 1$) de la ponderación entre términos del criterio (2), afectados por valores grandes de p_{ik} y los afectados por valores pequeños. Por otra parte, un valor de p_{ik} próximo a 1 (por ejemplo, $p_{ik} \geq 0.75$) indica que el objeto i se encuentra cerca de v_k, y un valor pequeño indicaría que i se encuentra cerca de la "frontera" de la k ésima clase difusa. O quizás, "entre" dos o más clases difusas. Así entonces, el exponente s actúa como un factor diferenciador de estas distintas situaciones.

La elaboración de un algoritmo para minimizar el criterio (2) se hace con base en las siguientes condiciones necesarias de minimización del criterio de clasificación:

1. Para una partición difusa fija P se minimiza el criterio (2) con respecto a V: $\text{Min}\{L(P,V,s) \,|v_k \in I\!R^h\}$. La solución del problema es la matriz $\widetilde{V}$ con filas $\widetilde{v}_k$ definidas por:

$$\widetilde{v}_k = \frac{\sum_{i=1}^{p} p_{ik}^s x_i}{\sum_{i=1}^{p} p_{ik}^s}; \; k = 1, \ldots, K. \tag{3}$$

2. Sea $s > 1$, V fija y M_{fr} el conjunto de las particiones difusas con K clases. Se minimiza el criterio (2) con respecto a $P = (p_{ik})$, sujeto a las p restricciones: $\sum_{k=1}^{K} p_{ik} = 1$ para todo i. Usando multiplicadores de Lagrange, se deducen condiciones necesarias para un mínimo local condicionado de $f(P) = L(P,V,s)$: $\widetilde{P} = (\widetilde{p}_{ik})$ donde:

$$\widetilde{p}_{ik} = \frac{\|x_i - v_k\|^{\frac{-2}{s-1}}}{\sum_{l=1}^{K} \|x_i - v_l\|^{\frac{-2}{s-1}}} \tag{4}$$

   siempre que para todo $i \in I$, $\|x_i - v_k\| > 0$ para todo k.

   En caso de que existan singularidades, es decir; existe $i \in I$ tal que $\|x_i - v_k\| = 0$ para algún k, entonces se define

$$\sum_{k \in I_i} \widetilde{p}_{ik} = 1 \;\text{ y }\; (\|x_i - v_k\| > 0 \iff \widetilde{p}_{ik} = 0) \tag{5}$$

   donde $I_i = \{k \in \{1, \ldots, K\} \,|\, \widetilde{p}_{ik} \neq 0\}$.

Observaciones

1. Sea $P \in M_{fr}$ que satisface la condición necesaria de mínimo condicionado de f, dada por (4). Y sea $\widetilde{P}$ la correspondiente partición difusa definida por las fórmulas (4) y (5). Entonces el racional de la fórmula (5) se explica porque al anularse los términos $\sum_{k=1}^{K} \widetilde{p}_{ik}^{s} \|x_i - v_k\|^2 = 0$ del criterio de clasificación donde ocurren singularidades, se tiene: $f(\widetilde{P}) \leq f(P)$.
2. La partición difusa $\widetilde{P}$ no es única, salvo si $I_i \neq \emptyset \Rightarrow |I_i| = 1$.

Si $s > 1$, un algoritmo "natural" para minimizar el criterio (2) es[1]:

1. Dar una partición difusa inicial P^0 (por ejemplo generada al azar).
2. Repetir (a) y (b) hasta que P^n se estabilice o se alcance un número máximo de iteraciones. Para $n = 0, 1, \ldots$
   (a) Usar P^n para calcular $V^n = (v_{kl}^n)$, de acuerdo con la fórmula (3).
   (b) Usar V^n para calcular P^{n+1}, de acuerdo con (4) y (5).

H.H. Bock probó, usando un argumento probabilístico, que la sucesión $L(P^n, V^n, s)$ es decreciente (ver [5]). Luego la sucesión converge, puesto que es inferiormente acotada. Es decir, el algoritmo converge.
Sea $\widetilde{V}$ definida de acuerdo con la fórmula (3). Se sabe que cada matriz P^* que minimiza la función $f : M_{fr} \to [0, +\infty[$ tal que $f(P) = L(P, \widetilde{V}, 1)$, es una partición (ver [5]). Este resultado dice que es posible obtener particiones óptimas, aplicando el algoritmo anterior con $s > 1$ y $s \approx 1$.
En lo que sigue de este documento, se formula un método de clasificación difusa con datos bimodales.

3 Un método de clasificación bimodal difusa

Sea $X = (x_{ij})_{p \times q}$ una matriz de similitudes entre objetos de de dos modos I y J, $I \cap J = \emptyset$, con $p = |I|$ y $q = |J|$. Los elementos de X reflejan la interacción o asociación entre los dos modos. Esto es, mientras más fuerte es la relación entre los objetos $i \in I$ y $j \in J$, más grande es x_{ij} y recíprocamente.
Los métodos para obtener clasificaciones del tipo $P \times Q$ donde P y Q son particiones de I y J con K y L clases respectivamente, se llaman métodos de particionamiento bimodal. Entre ellos se tienen los métodos inspirados en el modelo aditivo

$$\widehat{x}_{ij} = c + \sum_{k=1}^{K} \sum_{l=1}^{L} p_{ik} q_{jl} v_{kl}$$

[1]El algoritmo es conocido en la literatura especializada con el nombre de algoritmo ISODATA difuso (*fuzzy ISODATA algorithm*, en inglés).

donde el problema es hallar P, Q y $V = (v_{kl})_{K\times L}$ de modo que el criterio

$$L(P,Q,V) = \sum_{i=1}^{p}\sum_{j=1}^{q}(x_{ij} - \widehat{x}_{ij})^2$$

sea mínimo. Este criterio asume una forma simple:

$$L(P,Q,V) = \sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{i=1}^{p}\sum_{j=1}^{q} p_{ik}q_{jl}\,(x_{ij} - v_{kl})^2 \tag{6}$$

A partir de lo anterior se desarrollaron los métodos siguientes: intercambios alternantes de Gaul, W. y Schader, M. ([8]), un método de tipo k means de Baier, D.; Gaul, W. y Schader, M. ([1]) y un método de sobrecalentamiento simulado aplicado al esquema de intercambios alternantes propuesto por Trejos, J. y Castillo, W. ([13]) y Castillo, W. ([6]).

Para el caso de clases con intersecciones (*overlapping clustering*) también se han propuesto métodos para mimimizar el criterio (6), ver por ejemplo [1] y [8].

Con el propósito de formular un método de clasificación bimodal difusa definimos, en lo que sigue, un criterio de clasificación y un algoritmo para minimizarlo.

3.1 El criterio

Sea $s \geq 1$ y $F = (f_{ijkl})$. Se dice que F es una partición difusa bimodal si:

1. Para todo i, j, k, l; $f_{ijkl} \in [0,1]$.
2. Para todo $i \in I$ y $j \in J$; $\displaystyle\sum_{k=1}^{K}\sum_{l=1}^{L} f_{ijkl} = 1$.
3. Para todo k, l; $\displaystyle\sum_{i=1}^{p}\sum_{j=1}^{L} f_{ijkl} > 0$.

En esta definición f_{ijkl} es el grado de pertenencia de (i,j) a la clase bimodal difusa $A_k \times B_l$ determinada por un cierto umbral. Sea además, la matriz $V = (v_{kl})$ donde cada v_{kl} es representativo de la clase bimodal difusa $A_k \times B_l$. Se propone el siguiente criterio de clasificación bimodal difusa:

$$L(F,V,s) = \sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{i=1}^{p}\sum_{j=1}^{q} f_{ijkl}^{s}\,(x_{ij} - v_{kl})^2\,. \tag{7}$$

Puede notarse que el criterio (7) es una generalización del criterio (6), tomando $s = 1$ y $f_{ijkl} = p_{ik}q_{jl}$.

Con el fin de disminuir el número de parámetros por estimar se asumirá, en lo que sigue, una hipótesis adicional: $f_{ijkl} = p_{ik}q_{jl}$ donde, $P = (p_{ik})$ y $Q = (q_{jl})$ son particiones unimodales difusas de I y J con K y L clases, respectivamente. Esto es,

- Para todo $i, j, k, l; p_{ik}, q_{jl} \in [0,1]$.
- Para todo $i,\ j; \sum_{k=1}^{r} p_{ik} = 1 = \sum_{l=1}^{L} q_{jl}$.
- Para todo $k,\ l; \sum_{i=1}^{p} p_{ik} > 0$ y $\sum_{j=1}^{L} q_{jl} > 0$.

Los parámeros p_{ik} y q_{jl} vienen a ser los grados de pertenencia de i y j a las clases unimodales difusas A_k y B_l asociadas a las columnas k y l de P y Q respectivamente. En este sentido lo que se asume es similar a una hipótesis de "independencia" del grado de pertenencia.
El criterio (7) se escribe entonces como

$$L(P,Q,V,s) = \sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{i=1}^{p}\sum_{j=1}^{q} p_{ik}^{s} q_{jl}^{s} \left(x_{ij} - v_{kl}\right)^2. \tag{8}$$

3.2 Actualización de los promedios

La fórmula de actualización de V se deduce minimizando la función $\varphi(V) = L(P,Q,V,s)$, dadas las particiones difusas P y Q : sea $M_{K\times L}$ el conjunto de matrices $K \times L$ con entradas en $\mathbb{R}$, entonces, $\min\{\varphi(V) \mid V \in M_{K\times L}\}$ se alcanza en $\widetilde{V}$ definida por

$$\widetilde{v}_{kl} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q} p_{ik}^{s} q_{jl}^{s} x_{ij}}{\sum_{i=1}^{p}\sum_{j=1}^{q} p_{ik}^{s} q_{jl}^{s}}. \tag{9}$$

Este resultado se obtiene verificando que $\widetilde{V}$ es un punto estacionario de φ. Se nota que $\widetilde{V}$ es una matriz de promedios ponderados de las entradas x_{ij}.

3.3 Actualización de las particiones difusas

Las fórmulas de actualización de las particiones difusas P y Q se consiguen al identificar condiciones necesarias de mínimo para ciertas funciones.

Definición 1 *Sean* $d_{ik} = \sum_{j=1}^{q}\sum_{l=1}^{L} q_{jl}^{s}(x_{ij} - v_{kl})^2$ *con* $i \in I$ *y* $k \in \{1,\ldots,K\}$ *y* $e_{jl} = \sum_{i=1}^{p}\sum_{k=1}^{K} p_{ik}^{s}(x_{ij} - v_{kl})^2$ *con* $j \in J$ *y* $l \in \{1,\ldots,L\}$. *Se dice que* $i \in I$ ($j \in J$ *resp.*) *es una singularidad si* $d_{ik} = 0$ ($e_{jl} = 0$ *resp.*) *para algún* $k \in \{1,\ldots,K\}$ ($l \in \{1,\ldots,L\}$ *resp.*).

3.3.1 Actualización de P

Sean Q y V fijos, y $\phi(P) = L(P, Q, V, s)$ con $s > 1$. Se identificarán condiciones necesarias de mínimo condicionado del problema $\min_P \phi(P)$ sujeto a las p restricciones $\sum_{k=1}^{K} p_{ik} = 1$. El lagrangiano de la función ϕ es $\phi(P) + \sum_{i=1}^{p} \lambda_i \left[-1 + \sum_{k=1}^{K} p_{ik}\right]$.

Derivando parcialmente el lagrangiano con respecto a p_{ik} e igualando a cero su derivada, se llega a, $\lambda_i + s \sum_{j=1}^{q} \sum_{l=1}^{L} p_{ik}^{s-1} q_{jl}^{s} (x_{ij} - v_{kl})^2 = 0$. Sea $d_{ik} = \sum_{j=1}^{q} \sum_{l=1}^{L} q_{jl}^{s} (x_{ij} - v_{kl})^2$ y supongamos que i no es una singularidad, entonces despejando p_{ik} de la anterior ecuación se consigue,

$$p_{ik} = \left(\frac{-\lambda_i}{s}\right)^{\frac{1}{s-1}} d_{ik}^{\frac{-1}{s-1}}. \tag{10}$$

Luego, $\sum_h p_{ih} = 1 = \left(\frac{-\lambda_i}{s}\right)^{\frac{1}{s-1}} \sum_h d_{ih}^{\frac{-1}{s-1}}$ o sea $\left(\frac{-\lambda_i}{s}\right)^{\frac{1}{s-1}} = \frac{1}{\sum_h d_{ih}^{\frac{-1}{s-1}}}$. Finalmente, sustituyendo esto en (10) se tiene,

$$p_{ik} = \frac{d_{ik}^{\frac{-1}{s-1}}}{\sum_{h=1}^{K} d_{ih}^{\frac{-1}{s-1}}}. \tag{11}$$

Sea i una singularidad. Es decir, existe i, tal que $d_{it} = 0$ para algún t. Se define

$$\text{para todo } k \in I_i,\ p_{ik} > 0,\ \sum_{k \in I_i} p_{ik} = 1 \ \text{ y } \ (d_{ik} > 0 \iff p_{ik} = 0) \tag{12}$$

donde $I_i = \{k \in \{1, \ldots, K\} \,|\, d_{ik} = 0\}$.

Finalmente, es claro que P definida por las fórmulas (11) y (12) es una partición difusa. Además, no es única, salvo si $I_i \neq \emptyset \Rightarrow |I_i| = 1$.

3.3.2 Actualización de Q

Por un procedimiento análogo al anterior se deducen las fórmulas de actualización de la partición difusa Q.

Sea $e_{jl} = \sum_{i=1}^{p} \sum_{k=1}^{K} p_{ik}^{s} (x_{ij} - v_{kl})^2$ y supongamos que $e_{jl} > 0$ para todo l, entonces

$$q_{jl} = \frac{e_{jl}^{\frac{-1}{s-1}}}{\sum_{h=1}^{L} e_{jh}^{\frac{-1}{s-1}}}. \tag{13}$$

Si existe j, tal que $e_{jl} = 0$ para algún l, entonces se define

$$\text{para todo } j \in J_j,\ q_{jl} > 0,\ \sum_{l \in J_j} q_{jl} = 1 \ \text{ y } \ (e_{jl} > 0 \iff q_{jl} = 0) \tag{14}$$

donde $J_j = \{l \in \{1, \ldots, L\} \,|\, e_{jl} = 0\}$.

3.4 Minimización numérica del criterio $L(P, Q, V, s)$

Utilizando las fórmulas deducidas en las secciones anteriores se formula un algoritmo para minimizar el criterio $L(P, Q, V, s)$ con $s > 1$. Luego se prueba que este algoritmo converge.

3.4.1 El algoritmo

1. Dar las particiones difusas iniciales P^0 y Q^0 (por ejemplo generadas al azar). Calcular V^0 de acuerdo con la fórmula (9).

2. Repetir (a), (b) y (c) hasta que $1 - \frac{L_n}{L_0} > umb$, donde $L_n = L(P^n, Q^n, V^n, s)$ y $umb \in [\alpha, 1[$ con $\alpha \approx 1$; o se alcance un número máximo de iteraciones.

   Para $n = 0, 1, \ldots$:

   (a) Usar Q^n y V^n para calcular $P^{n+1} = (p_{ik}^{n+1})$, de acuerdo con (11) y (12).

   (b) Usar P^{n+1} y V^n para calcular $Q^{n+1} = (q_{jl}^{n+1})$, de acuerdo con (13) y (14).

   (c) Usar P^{n+1} y Q^{n+1} para calcular $V^{n+1} = (v_{kl}^{n+1})$, de acuerdo con (9).

3.4.2 Convergencia del algoritmo

Proposición 1 *Sea $L_n = L(P^n, Q^n, V^n, s)$. El algoritmo recién definido tiene las siguientes propiedades:*

(a) Para todo $n = 0, 1, \ldots$; $L_n \geq L_{n+1}$. Por lo tanto la sucesión (L_n) converge.

(b) Sea n tal que $L_n = L_{n+1}$, entonces para todo i y j que no son singularidades se tiene: $p_{ik}^n = p_{ik}^{n+1}$ y $q_{jl}^n = q_{jl}^{n+1}$ para todo $k \in \{1, \ldots, K\}$ y $l \in \{1, \ldots, L\}$.

DEMOSTRACIÓN: La propiedad (a) se obtiene probando las desigualdades

$$L_n \geq L(P^{n+1}, Q^n, V^n, s) \geq L(P^{n+1}, Q^{n+1}, V^n, s) \geq L_{n+1}.$$

para lo cual se usará un argumento probabilístico (ver [5]). Primero se demostrará que $\forall\ n$, $L_n \geq L(P^{n+1}, Q^n, V^n, s)$. Es suficiente probar, para todo $i \in I$,

$$\sum_{k=1}^{K} (p_{ik}^n)^s\, d_{ik}^n \geq \sum_{k=1}^{K} \left(p_{ik}^{n+1}\right)^s d_{ik}^n$$

donde $d_{ik}^n = \sum_{j=1}^{q} \sum_{l=1}^{L} \left(q_{jl}^n\right)^s (x_{ij} - v_{kl}^n)^2$. Si i es una singularidad, entonces $0 = \sum_{k=1}^{K} \left(p_{ik}^{n+1}\right)^s d_{ik}^n$ y vale la desigualdad anterior para este i. En cualquier otro caso, sustituyendo $p_{ik}^{n+1} = \frac{\left(d_{ik}^n\right)^{\frac{-1}{s-1}}}{\sum_{h=1}^{K} \left(d_{ik}^n\right)^{\frac{-1}{s-1}}}$, en el lado derecho de la desigualdad anterior, vemos

que éste es igual a $\left[\sum_{k=1}^{K}(d_{ik}^n)^{\frac{-1}{s-1}}\right]^{1-s}$. Por lo tanto es suficiente probar que, para todo i,

$$\sum_{k=1}^{K}(p_{ik}^n)^s\, d_{ik}^n \geq \left[\sum_{k=1}^{K}(d_{ik}^n)^{\frac{-1}{s-1}}\right]^{1-s}. \tag{15}$$

Sea para todo k, $y_k = (p_{ik}^n)^{s-1}d_{ik}^n$. Se define la variable aleatoria Y tal que para todo k, $\Pr(Y = y_k) = p_{ik}^n$. Teniendo en cuenta que la función $h(y) = y^{\frac{-1}{s-1}}$ es convexa y estrictamente decreciente para $y > 0$, resulta fácil deducir lo siguiente:

1. $E[Y] = \sum_{k=1}^{K}(p_{ik}^n)^s\, d_{ik}^n$.
2. $E[h(Y)] = \sum_{k=1}^{K}(d_{ik}^n)^{\frac{-1}{s-1}} = h\left([\sum_{k=1}^{K}(d_{ik}^n)^{\frac{-1}{s-1}}]^{1-s}\right)$.

Por la desigualdad de Jensen se tiene $h(E[Y]) \leq E[h(Y)]$. Es decir,

$$h(\sum_{k=1}^{K}(p_{ik}^n)^s\, d_{ik}^n) \leq h\left(\left[\sum_{k=1}^{K}(d_{ik}^n)^{\frac{-1}{s-1}}\right]^{1-s}\right).$$

Como h es estrictamente decreciente y continua (y por tanto biyectiva), en $]0,+\infty[$, entonces,

$$\sum_{k=1}^{K}(p_{ik}^n)^s\, d_{ik}^n \geq \left[\sum_{k=1}^{K}(d_{ik}^n)^{\frac{-1}{s-1}}\right]^{1-s}$$

que es la desigualdad (15).

La desigualdad $L(P^{n+1}, Q^n, V^n, s) \geq L(P^{n+1}, Q^{n+1}, V^n, s)$ se prueba de manera similar. Finalmente, $L(P^{n+1}, Q^{n+1}, V^n, s) \geq L_{n+1}$ es consecuencia de $L_{n+1} = \min_V L(P^{n+1}, Q^{n+1}, V, s)$.

Para demostrar la propiedad (b), se observa en particular que $L_n = L(P^(n+1), Q^n, V^n, s)$ es consecuencia de la hipótesis $L_n = L_{n+1}$. En virtud de esto y de la desigualdad (15), se obtiene para todo i que no es una singularidad,la igualdad

$$\sum_{k=1}^{K}(p_{ik}^n)^s\, d_{ik}^n = \left[\sum_{k=1}^{K}(d_{ik}^n)^{\frac{-1}{s-1}}\right]^{1-s}.$$

Es decir, $h(E[Y]) = E[h(Y)]$. En consecuencia la variable aleatoria Y es constante[2]. Esto es, para todo k, $(p_{ik}^n)^{s-1}\, d_{ik}^n = c > 0$. O sea, $p_{ik}^n = c^{\frac{1}{s-1}}\,(d_{ik}^n)^{\frac{-1}{s-1}}$.

[2]En el caso general, $Y = c$ con probabilidad 1. En nuestro caso particular, h es estrictamente convexa y derivable en $E[Y] \in]0,+\infty[$. Al considerar la recta tangente a la gráfica de h por el punto $(E[Y], h(E[Y]))$ se puede deducir que Y es constante si y solo si $E[h(Y)] = h(E[Y])$.

Por otra parte, de acuerdo con (11), $p_{ik}^{n+1} = \frac{(d_{ik}^n)^{\frac{-1}{s-1}}}{\sum_{l=1}^K (d_{il}^n)^{\frac{-1}{s-1}}} = \alpha p_{ik}^n$ con $\alpha = \left(\sum_{l=1}^K (d_{il}^n)^{\frac{-1}{s-1}}\right)^{-1} c^{\frac{-1}{s-1}}$. Como $\sum_{k=1}^K p_{ik}^n = 1$ entonces $\alpha = 1$. ∎

3.5 Análisis del caso $s = 1$

Sea la función $\psi(P) = \min_V L(P, Q, V, 1)$ con Q fija. Identificamos la matriz P de tamaño $p \times K$ con el vector $(p_1, \ldots, p_p) \in I\!R^{pr}$, donde p_i es la fila i de la matriz P. Sean S y T los conjuntos definidos por[3] $S = \left\{ P \in I\!R^{pr} \;\middle|\; \text{para todo } i,\ p_i \geq 0 \text{ y } \sum_{k=1}^K p_{ik} = 1 \right\}$ y $T = \left\{ Q \in I\!R^{qm} \;\middle|\; \text{para todo } j,\ q_j \geq 0 \text{ y } \sum_{l=1}^L q_{jl} = 1 \right\}$. En lo que sigue se prueba que la función ψ alcanza su mínimo en una partición $P^* \in S$ (proposición 5) y que si $\Phi(P, Q) = \min_V L(P, Q, V, 1)$ alcanza su mínimo en $(P^*, Q^*) \in S \times T$, entonces P^* y Q^* son particiones (ver proposición 6).

Proposición 2 *S es un poliedro[4] acotado.*

DEMOSTRACIÓN: Sean 1_p y 0_{pr} las columnas de unos y ceros, de longitud p y pr respectivamente. S es la intersección de los poliedros $\{x \in I\!R^{pr} \,|Ax \leq 1_p\}$, $\{x \in I\!R^{pr} \,|-Ax \leq -1_p\}$ y $\{x \in I\!R^{pr} \,|-Bx \leq 0_{pr}\}$ donde A y B son, respectivamente, las matrices de coeficientes de los sistemas $\sum_{k=1}^K p_{ik} = 1$, $i = 1, \ldots, p$ y $p_{ik} = 0$, $i = 1, \ldots, p$, $k = 1, \ldots, K$. Luego S es un poliedro.
Por otra parte, es claro que $S \subseteq [0, 1]^{pr}$. Luego S es acotado. ∎

Proposición 3 *P es un extremo[5] de S si y sólo si P es una partición.*

DEMOSTRACIÓN: Sea $P \in S$ una partición que no es un extremo de S. Razonemos por contradicción. Existen $P' \neq P''$ y $\alpha \in]0, 1[$ tales que $P = \alpha P' + (1-\alpha)P''$ con P', $P'' \in S$. Entonces existe k' tal que para todo $k \neq k'$, $p_{ik} = \alpha p'_{ik} + (1 - \alpha)p''_{ik} = 0$; luego para todo $k \neq k'$, $p'_{ik} = p''_{ik} = 0$ y $p'_{ik'} = p''_{ik'} = 1$. Es decir $P' = P''$, lo cual es una contradicción.
Por otro lado, sea $P \in S$. Se prueba que si P no es una partición entonces P no es un extremo. En efecto, como $P = (p_{ik})$ no es una partición, existen i y $K_i \subset \{1, \ldots, K\}$ tales que

- $|K_i| \geq 2$, $p_{ik} \in]0, 1[$ para todo $k \in K_i$, y
- si $K_i \neq \{1, \ldots, K\}$ entonces $p_{ik} = 0$ para todo $i \notin K_i$.

---

[3] Esta definición admite posibles clases difusas vacías para $P \in S$.

[4] Un poliedro en $I\!R^n$ es la intersección de un número finito de semiespacios cerrados de $I\!R^n$. Un semiespacio en $I\!R^n$ es un conjunto de la forma $\{x \in I\!R^n \,|Ax \leq b\}$ donde A y b son matrices fijas $p \times n$ y $p \times 1$ respectivamente.

[5] Por definición $P \in S$ es un extremo del convexo S si no existen $P' \neq P''$ y $\alpha \in]0, 1[$ tales que $P = \alpha P' + (1 - \alpha)P''$ con P', $P'' \in S$.

Es claro que $P = \sum_{k\in K_i} p_{ik}P^{(k)}$ donde $P^{(k)} = \left(p_{is}^{(k)}\right)_{p\times K}$ es igual a P excepto por las entradas $p_{is}^{(k)}$, que se definen así:

$$p_{is}^{(k)} = \begin{cases} 1 & \text{si } s = k. \\ 0 & \text{en caso contrario.} \end{cases}$$

Sean $\alpha = \sum_{k\in K_i-\{k'\}} p_{ik}$ y $\overline{P} = \sum_{k\in K_i-\{k'\}} \frac{p_{ik}}{\alpha} P^{(k)} \in S$. Es claro que $P = p_{ik'}P^{(k')} + \alpha\overline{P}$ con $\{p_{ik'}, \alpha\} \subset]0,1[$ y $p_{ik'} + \alpha = 1$. Además, $P^{(k')} \neq \overline{P}$, puesto que $p_{ik'}^{(k')} = 1$ y $\left(\overline{P}\right)_{ik'} = \sum_{k\in K_i-\{k'\}} \frac{p_{ik}}{\alpha} p_{ik'}^{(k)} = 0$. Luego P no es un extremo de S. ∎

Proposición 4 *ψ es cóncava en S.*

DEMOSTRACIÓN: sea $\alpha \in [0,1]$ y P', $P'' \in S$, entonces

$$\begin{aligned}
\psi(\alpha P' + (1-\alpha)P'') &= \min_V \sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{i=1}^{p}\sum_{j=1}^{q} \left(\alpha p'_{ik} + (1-\alpha)\, p''_{ik}\right) q_{jl}\left(x_{ij} - v_{kl}\right)^2 \\
&= \sum_{k=1}^{K}\sum_{l=1}^{L} \min_{v_{kl}} \left[\sum_{i=1}^{p}\sum_{j=1}^{q} \alpha p'_{ik} q_{jl}\left(x_{ij} - v_{kl}\right)^2 \right. \\
&\quad \left. + \sum_{i=1}^{p}\sum_{j=1}^{q} (1-\alpha)\, p''_{ik} q_{jl}\left(x_{ij} - v_{kl}\right)^2\right] \\
&\geq \alpha \sum_{k=1}^{K}\sum_{l=1}^{L} \min_{v_{kl}} \sum_{i=1}^{p}\sum_{j=1}^{q} p'_{ik} q_{jl}\left(x_{ij} - v_{kl}\right)^2 \\
&\quad + (1-\alpha) \sum_{k=1}^{K}\sum_{l=1}^{L} \min_{v_{kl}} \sum_{i=1}^{p}\sum_{j=1}^{q} p''_{ik} q_{jl}\left(x_{ij} - v_{kl}\right)^2 \\
&= \alpha\psi(P') + (1-\alpha)\psi(P''). \blacksquare
\end{aligned}$$

Proposición 5 *Existe una partición $P^* \in S$ tal que $\psi(P^*) = \min_{P\in S} \psi(P)$.*

DEMOSTRACIÓN: La función ψ es cóncava (por proposición 4) y acotada inferiormente sobre el poliedro acotado S. Luego ψ alcanza su mínimo global en $P^* \in S$ y P^* es un punto extremo de S (ver [12])[6]. Por la proposición 3, P^*es una partición. ∎

Proposición 6 *Si la función $\Phi(P,Q) = \min_V L(P,Q,V,1)$ alcanza su mínimo en $(P^*,Q^*) \in S\times T$, entonces P^* y Q^* son particiones.*

[6]El teorema que se está aplicando aparece en la página 125 del libro de Roberts, A.W. y Varberg D.E. ([12]), como "Theorem E".

DEMOSTRACIÓN: Sea $\psi(P) = \min_V L(P, Q^*, V, 1)$ entonces

$$\psi(P^*) = \Phi(P^*, Q^*) \leq \Phi(P, Q^*) = \psi(P)$$

para todo $P \in S$. Luego P^* es una partición (por la proposición 5).
De manera similar, si $\varphi(Q) = \min_V L(P^*, Q, V, 1)$ entonces $\varphi(Q^*) = \Phi(P^*, Q^*) \leq \Phi(P^*, Q) = \varphi(Q)$ para todo $Q \in T$. Luego Q^* es una partición. ∎

4 Algunos resultados comparativos

Con el propósito de investigar la eficacia del nuevo método difuso (MD), éste fue programado con *Mathematica* y evaluado a partir de cálculos efectuados con datos que han sido objeto de investigación con otros métodos. De acuerdo con la proposición 6 demostrado en la sección anterior, escogiendo $s \approx 1$, el MD puede ser usado para detectar particiones eventualmente óptimas. Con esta idea se abordó la presente investigación comparativa.

El conjunto de datos usado corresponde a una tabla de ochenta y dos filas y veinticinco columnas y se ubica en un estudio de mercado para introducir un nuevo producto de la línea de sanitarios ecológicos (ver [1])[7]. Cada entrada $x_{ij} \in \{1, \ldots, 11\}$ de esta tabla expresa la intención de compra del cliente potencial i, con base en el nivel j de un atributo del producto. Se tomaron en cuenta cinco atributos (como precio, reciclado, etc.) cada uno con cinco niveles.

La siguiente tabla contiene el mejor valor encontrado de la varianza explicada VE definida por $VE = 1 - \frac{\sum_{k=1}^{K}\sum_{l=1}^{L} p_{ik} q_{jl} (x_{ij} - \widehat{x}_{ij})^2}{\sum_{i=1}^{p}\sum_{j=1}^{q} (x_{ij} - \overline{x})^2}$ donde $\overline{x} = \frac{1}{pq}\sum_{i=1}^{p}\sum_{j=1}^{q} x_{ij}$. Se comparan el método de tipo k-means, el de intercambios alternantes (IA) y MD.

$K = L$	k-means	IA	MD
3	0.3234	0.3234	0.3559
4	0.3729	0.3729	0.4000

En el caso del método MD se usó $s = 1.01$.

En [1] se reportan los mismos valores de VE como los mejores valores encontrados por el método k-means en cincuenta inicios al azar. Los valores de VE obtenidos con el método MD son significativamente mayores que los obtenidos con los otros métodos, no obstante que el criterio opti optimizado por MD no es el de la varianza explicada.

El método MD también fue aplicado con valores pequeños de s a unos datos sobre bebidas gaseosas, obteniéndose la misma partición en tres clases de Eckes & Orlick ([7]). Con los datos Cigarrilos 1, Cigarrillos 2 y Coñac, se encontraron los mismos resultados conocidos y reportados en [6] y [8], para $K = L = 3$ y $K = L = 4$.

Se realizaron pruebas para valores grandes de s ($s = 1.5, 2$, por ejemplo) con los datos de sanitarios y los datos de *conductas versus situaciones* analizados en [7] y en [11]. Sin embargo el algoritmo falló en la detección de clases difusas.

[7] Agradecemos a los autores de este artículo por enviarnos los datos

Actualmente, estamos realizando experimentaciones por medio de simulaciones de Monte Carlo [10] para evalkuar el método con una variante que hace disminuir el valor de s, y se hacen comparaciones con otros cuatro métodos.

Referencias

[1] Baier, D.; Gaul, W.; Schader, M. (1997) "Two-mode overlapping clustering with applications to simultaneous benefit segmentation and market structuring", in: R. Klar & O. Opitz (Eds.), *Classification and Knowledge Organization*. Springer, Heidelberg, 557–566.

[2] Bezdek, J.C. (1980) "A convergence theorem for the fuzzy ISODATA clustering algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(1): 1–8.

[3] Bezdek, J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York and London.

[4] Bezdek, J.C.; Hathaway, R.J.; Sabin, M.J.; Tucker, W.T. (1987) "Convergence theory for fuzzy c-means: counterexamples and repairs", *IEEE Transactions on Systems, Man and Cybernetics*, 17(5): 873–877.

[5] Bock, H.H. (1979) "Fuzzy clustering procedures", in: R. Tomassone (Ed.) *Analyse des Données et Informatique*, INRIA, Paris: 205–218.

[6] Castillo, W. (1999) *Métodos de Particionamiento Bimodal y Trimodal*. Tesis de Maestría, Universidad de Costa Rica.

[7] Eckes, T.; Orlik, P. (1993) "An error variance approach to two-mode hierarchical clustering", *Journal of Classification* 10, 51-74.

[8] Gaul, W.; Schader, M. (1996) "A new algorithm for two-mode clustering", in: H.H. Bock & W. Polasek (Eds.) *Data Analysis and Information Systems*. Springer, Heidelberg, 15–23.

[9] Groenen, P.J.F.; Jajuga, K. (2001) "Fuzzy clustering with squared Minkowski distances", *Fuzzy Sets and Systems* **120**(2): 227–237.

[10] Groenen, P.J.F.; Rosmalen, J. van; Trejos, J.; Castillo, W. (2009) "Optimization strategies for two-mode partitioning", aceptado en *Journal of Classification*.

[11] Mirkin, B.G.; Arabie, P.; Hubert, L.J. (1995) "Additive two-mode clustering: the error-variance approach revisited", *Journal of Classification* **12**(2): 243–263.

[12] Roberts, A.W.; Varberg, D.E. (1973) *Convex Functions*. Academic Press, New York.

[13] Trejos, J.; Castillo, W. (2000) "Simulated annealing optimization for two-mode partitioning", in: W. Gaul & R. Decker (Eds.) *Classification and Information at the Turn of the Millenium*. Springer-Verlag, Berlin: 133–142.