

# PARTICIONES ÓPTIMAS: CARACTERÍSTICAS Y CALIDAD DE SUS APROXIMACIONES

SAID LABRÈCHE\*

---

## Resumen

En la primera parte, se hará una presentación factorial del problema de clasificación según el criterio de mínimos cuadrados, para toda escogencia de la métrica en el espacio de individuos. Se deduce que la inercia interclases posee una cota superior que depende del número de clases y de los resultados de un Análisis en Componentes Principales, lo que nos permite generalizar un coeficiente para medir la calidad de la aproximación de una partición óptima.

En la segunda parte, se da una demostración original del hecho que la inercia induce un orden estricto hasta un cierto rango sobre cualquier conjunto de particiones óptimas. Finalmente, mediante un procedimiento heurístico se propone una manera de escoger *a priori* el número de clases en una población.

**Palabras clave:** análisis en componentes principales, análisis factorial, clasificación automática, clasificación por particiones, partición óptima, número de clases.

## Abstract

In the part, we present a factorial approach for clustering following the least squares criterion, for every choice of the metrics in the individual space. We deduce that the between-clusters inertia has an upper bound that depends on the number of clusters and the results of a Principal Component Analysis; this enables us to generalize a coefficient that measures the quality of the approximation of an optimal partition.

In the second part, we demonstrate that the inertia induces a strict ordering of the set of optimal partitions. Finally, we propose a heuristic for choosing the number of clusters.

## 1. Introducción

Muchos métodos de Clasificación Automática no jerárquica consisten en buscar una partición de una población en un cierto número de clases, de manera tal que se optimice un

---

\*Universidad Paul Sabatier, Toulouse, Francia

criterio matemático bien definido [7, 8, 15, 16, 20]. Uno de estos criterios es el de la inercia interclases [5, 9], también llamado criterio de la varianza debido a su significado estadístico [12, 17]. Será de este criterio que trataremos en este artículo.

La presentación factorial del problema de la clasificación ha sido tratada por diversos autores [2, 13, 16, 17, 18, 21]. Por ejemplo, Lerman se restringe al caso en que la matriz asociada a la métrica escogida en el espacio de los individuos es diagonal; nuestro trabajo es una extensión al caso general de esta presentación. Deduciremos también que la inercia explicada por toda partición tiene una cota superior que es menor o igual a la inercia total. Esta cota depende del número de clases y de los resultados de un Análisis en Componentes Principales (ACP) de la nube de puntos a clasificar.

Este resultado nos permitirá proponer un coeficiente para controlar la calidad de la aproximación de una partición óptima por toda partición en un número dado de clases.

Daremos una demostración original del hecho que la inercia interclases induce un orden estricto hasta un cierto rango sobre todo conjunto de particiones óptimas.

A partir de estos dos resultados y de los resultados significativos obtenidos sobre varias tablas de datos, reales o simuladas, proponemos un método que permite resolver parcialmente el problema abierto [5, 7, 8, 11] de la escogencia del número de clases.

## 2. Presentación factorial de la clasificación

### 2.1. Datos y notaciones

Los datos son presentados bajo la forma de una matriz *individuos*  $\times$  *variables*  $X$  de dimensiones  $n \times p$  y de rango  $s$ . Esta matriz representa las observaciones de  $p$  variables centradas sobre un conjunto  $\Omega$  de  $n$  individuos. La entrada  $(i, j)$  de  $X$  es el valor que toma la  $j$ -ésima variable sobre el  $i$ -ésimo individuo.

Sea  $M$  la matriz de la métrica escogida en el espacio  $E$  de individuos y  $D$  la matriz diagonal de los pesos de los individuos que define la métrica en el espacio  $F$  de variables. El triplete  $(X, M, D)$  define una nube de puntos  $\mathcal{N}_x$  en  $E$ . El  $i$ -ésimo punto  $X_i$  de  $\mathcal{N}_x$  representa la  $i$ -ésima fila de  $X$ .

La inercia interclases de una partición  $P_q = \{P_q(k)/k = 1, \dots, q\}$  de  $\Omega$  en  $q$  clases está definida por:

$$B(P_q) = \sum_{k=1}^q B[P_q(k)] = \sum_{k=1}^q \mu(k) \|G_k\|_M^2$$

donde:  $B[P_q(k)] = \mu(k) \|G_k\|_M^2$  es la contribución de la clase  $P_q(k)$  a  $B(P_q)$ ,  
 $G_k$  (resp.  $\mu(k)$ ) es el centro de gravedad (resp. peso) de la clase  $P_q(k)$ .

Si denotamos  $Y$  la matriz  $n \times q$  de las indicatrices asociadas a las clases de la partición  $P_q$ , tenemos una nueva expresión de la inercia interclases:

$$\begin{aligned} B(P_q) &= \text{traza}[(Y^T D Y)^{-1} Y^T D X M X^T D Y] \\ &= \text{traza}[\chi^2 V_{yx} M V_{xy}] \end{aligned}$$

donde  $\chi^2$  es la matriz diagonal del chi-cuadrado para las indicatrices (es decir,

$\chi^2(k, k) = 1/\mu(k)$  con  $\mu(k)$  el peso de la  $k$ -ésima clase) y  $V_{xy} (= X^T D Y = V_{yx}^T)$  es la matriz de covarianzas entre las variables  $\{X^j/j = 1, \dots, p\}$  y las indicatrices de las clases.

El problema de la clasificación consiste en buscar en el conjunto  $\mathcal{P}_q$  de todas las particiones de  $\Omega$  en  $q$  clases, una partición  $P_q^*$  que maximice  $B(P_q)$ . Tales particiones  $P_q^*$  serán llamadas *particiones óptimas*. Debe notarse que este problema puede tener varias soluciones distintas.

## 2.2. Definición de una nube de puntos en $F$ .

Como la matriz  $M$  es simétrica, definida, positiva y de rango  $p$ , existe una matriz  $M_1$  de rango  $p$  tal que:  $M = M_1 M_1^T$ . ( $M_1$  puede ser obtenida ya sea mediante una descomposición espectral de  $M$  o mediante la descomposición de Doolittle-Cholesky [3]).

Sean  $Z = X M_1$  y  $\mathcal{N}_z = \{Z^j/j = 1, \dots, p\}$  la nube de puntos en  $F$  asociados a los vectores columna de  $Z$ . Ponderamos cada punto de  $\mathcal{N}_z$  por  $1/p$  y denotamos  $D_p$  la matriz de pesos asociada.

### Observaciones:

- a) La nube  $\mathcal{N}_z$  está en el subespacio  $F_x$  de  $F$  generado por los vectores columna de  $X$ .
- b) La inercia de las nubes  $\mathcal{N}_x$  y  $\mathcal{N}_z$  son proporcionales. Verifican las siguientes igualdades:

$$I[\mathcal{N}_x] = \sum_{i=1}^n \mu_i \|X_i\|_M^2 = p \sum_{j=1}^p \frac{1}{p} \|Z^j\|_D^2 = p I[\mathcal{N}_z]$$

Sea  $R$  la matriz cuadrada  $Z D_p Z^T D$ , de rango  $s$ . Su  $j$ -ésimo valor propio no nulo  $\lambda_z^j$  es igual, salvo por un coeficiente de  $1/p$ , al  $j$ -ésimo momento principal  $\lambda_x^j$  del ACP de  $(\mathcal{N}_x, M, D)$ . Las variables principales asociadas a  $\lambda_x^j$  son vectores propios de  $R$  asociados a  $\lambda_z^j$ .

## 2.3. Proyección de $\mathcal{N}_z$ sobre $F_y$

En esta sección mostraremos que  $B(P_q)$  es proporcional a la inercia de la nube  $\mathcal{N}_z$  proyectada sobre el subespacio  $F_{yc}$  de dimensión  $q - 1$  generado por las indicatrices centradas de las clases. Denotamos  $Y_c$  la matriz de las indicatrices centradas de las clases. El subespacio  $F_y$  generado por las indicatrices (no centradas) se descompone como sigue:  $F_y = \Delta_{1n} \oplus F_{yc}$  donde  $\Delta_{1n}$  es el eje de constantes generado por el vector  $1_n$  cuyas componentes son todas iguales a 1.  $\Delta_{1n}$  es  $D$ -ortogonal a  $F_{yc}$  y  $F_x$ . El operador  $\Pi_y$  de proyección sobre  $F_y$  es la suma de los operadores de proyección  $\Pi_{yc}$  y  $\Pi_{1n}$  sobre  $F_{yc}$  y  $\Delta_{1n}$  respectivamente. Como  $\mathcal{N}_z$  está contenido en  $F_x$ , se deduce que  $\hat{\mathcal{N}}_z = \Pi_y(\mathcal{N}_z)$ , la imagen de  $\mathcal{N}_z$  por  $\Pi_y$ , está contenida en  $F_{yc}$ .

Considerando las expresiones de los operadores de proyección  $\Pi_y = Y(Y^T D Y)^{-1} Y^T D$  y  $\Pi_{yc} = Y_c(Y_c^T D Y_c)^\dagger Y_c^T D$ , donde  $\dagger$  simboliza la inversa generalizada de Moore-Penrose [22],

y teniendo en cuenta las propiedades de la traza de una matriz, la inercia de  $\widehat{\mathcal{N}}_z$  se puede escribir:

$$\begin{aligned}
I[\widehat{\mathcal{N}}_z] &= \sum_{j=1}^p (1/p) \|\Pi_y(Z^j)\|_D^2 = (1/p) \sum_{j=1}^p Y(Y^T D Y)^{-1} Y^T D Z^j \\
&= (1/p) \text{traza}[Y(Y^T D Y)^{-1} Y^T D X M X^T D] = (1/p) \text{traza}[\chi^2 V_{yx} M V_{xy}] \\
&= (1/p) B(P_q) \\
&= (1/p) \sum_{j=1}^p \|\Pi_{y_c}(Z^j)\|_D^2 = (1/p) \sum_{j=1}^p \|Y_c(Y^T D Y_c)^\dagger Y_c^T D Z^j\|_D^2 \\
&= (1/p) \text{traza}[Y_c(Y_c^T D Y_c)^\dagger Y_c^T D X M X^T D] \\
&= (1/p) \text{traza}[V_y^\dagger V_{yx} M V_{xy}] \\
&= (1/p) B(P_q)
\end{aligned}$$

**Consecuencia:**

Si  $\{U_k/k = 1, \dots, q-1\}$  es una base  $D$ -ortonormada de  $F_{y_c}$ , tenemos entonces una nueva expresión de  $I[\widehat{\mathcal{N}}_z]$ :

$$\begin{aligned}
I[\widehat{\mathcal{N}}_z] &= \sum_{j=1}^p \left\| \sum_{k=1}^{q-1} U_k U_k^T D Z^j \right\|_D^2 = \sum_{k=1}^{q-1} U_k^T D Z D_p Z^T D U_k \\
&= \sum_{k=1}^{q-1} U_k^T D R U_k = \sum_{k=1}^{q-1} D(R U_k, U_k) = (1/p) B(P_q).
\end{aligned}$$

Por lo tanto, buscar una partición del conjunto de  $\Omega$  que maximice  $B(P_q)$ , se reduce a buscar  $q-1$  vectores particulares  $D$ -ortonormados del espacio  $F$ , tales que la inercia de la nube  $\mathcal{N}_z$  proyectada sobre el subespacio que ellos generan sea máxima. Estos vectores deben generar un subespacio de  $F$  asociado al conjunto de indicatrices centradas de una partición en  $q$  clases.

La clasificación puede entonces ser presentada como un ACP de la nube  $(\mathcal{N}_z, D, D_p)$  con restricciones sobre los vectores axiales principales.

**Observaciones:**

- Cuando  $M$  es una matriz diagonal (resp. identidad), la presentación anterior es equivalente a la presentación de Lerman [17] (resp. Mirkin [21]).
- Con la introducción de las métricas a efecto relacional [24], también se muestra que la clasificación puede ser presentada como un ACP de la nube  $\mathcal{N}_x$ , con restricciones sobre los vectores axiales principales [16].

## 2.4. Cota superior de la inercia interclases

Antes de mostrar una propiedad extremal de  $B(p)$ , recordamos un resultado propuesto por Rao [23, p.331]:

sean  $A$  una matriz simétrica  $n \times n$  y  $D$  una matriz simétrica, definida positiva  $n \times n$ , si  $\{U_k/k = 1, \dots, q\}$  es un conjunto de vectores linealmente independientes de  $\mathbb{R}^n$  tales que:

$$\forall(k, k') \quad B(U_k, U_{k'}) = {}^tU_k D U_{k'} = \delta_{kk'} \quad (1)$$

donde  $\delta$  denota la delta de Kronecker, tenemos:

$$\sum_{k=1}^q {}^tU_k A U_k \leq \sum_{k=1}^q \lambda_k \quad (2)$$

donde  $\lambda_k$  es la  $k$ -ésima raíz de la ecuación  $\det(A - \lambda B) = 0$ ,  $\lambda \in \mathbb{R}$  y  $\det$  denota el determinante.

### Propiedad 1

La inercia interclases explicada por toda partición  $P_q$  en  $q$  clases cumple:

$$B(P_q) \leq \sum_{j=1}^{s_q} \lambda_x^j$$

donde  $s_q$  es el mínimo entre  $q - 1$  y  $s$ , el rango de  $X$ .

DEMOSTRACIÓN:

En las igualdades

$$I[\widehat{N}_z] = \sum_{k=1}^{q-1} {}^tU_k D R U_k = \sum_{k=1}^{q-1} D(RU_k, U_k) = \frac{1}{p} B(P)$$

la matriz  $DR$  es simétrica y positiva.

Sea  $\{\lambda_z^j/j = 1, \dots, s\}$  el conjunto de raíces no nulas de la ecuación

$\det(DR - \lambda D) = \det(D)\det(R - \lambda I_n)$ , donde  $I_n$  denota la matriz identidad de orden  $n$ .

Como los vectores  $\{U_k/k = 1, \dots, q - 1\}$  verifican la ecuación 1, de la desigualdad 2 deducimos

$$B(P) = pI[\widehat{N}_z] = p \sum_{k=1}^{q-1} {}^tU_k A U_k \leq p \sum_{j=1}^{s_q} \lambda_z^j = \sum_{j=1}^{s_q} \lambda_x^j$$

La propiedad queda así demostrada. ■

### Observaciones:

Una propiedad similar fue propuesta por F. Marchotorchino en el marco del Análisis Factorial Relacional, que el autor introdujo [19, 20]. Su demostración está basada en el teorema de Hoffman-Wielandt [19, pp. 87-88].

## 2.5. Proposición para medir la calidad de la aproximación

El problema de buscar las particiones óptimas es un problema combinatorio pues el cardinal de todo conjunto  $\mathcal{P}_q$  es finito [6, p.81]. En la práctica, estas particiones son difíciles de obtener y por lo tanto deben ser aproximadas mediante algoritmos. Una literatura abundante es consagrada a los métodos de aproximación, llamados algoritmos de clasificación [5, 6, 7, 8].

En los diferentes artículos y obras que hemos consultado y que tratan sobre la clasificación, no hemos encontrado métodos para “medir” globalmente la calidad de la aproximación de una solución exacta, de las soluciones obtenidas mediante los algoritmos. Por ello proponemos un coeficiente que cumpla esta función.

De la propiedad 1 deducimos que, para toda partición  $P_q$  en  $q$  clases, el cociente  $\frac{B(P_q)}{B(P_q^*)}$  está minorado por  $\frac{B(P_q)}{s_q}$ .

$$\sum_{j=1}^j \lambda_x^j$$

Para medir la calidad de la aproximación de una partición óptima  $P_q^*$  por  $P_q$ , proponemos el coeficiente  $QA$  (Calidad de la Aproximación) definido por:

$$QA(P_q) = 1 - \frac{B(P_q)}{s_q} \sum_{j=1}^j \lambda_x^j$$

Diremos entonces que la partición  $P_q$  es una **buena aproximación** de una partición óptima  $P_q^*$  si  $QA(P_q)$  está cercano a 0.

### Observaciones:

- Si el número  $q$  de clases es mayor o igual a  $s + 1$ , se tiene entonces  $\sum_{j=1}^{s_q} \lambda_x^j = I[\mathcal{N}_x]$ .

En este caso, el coeficiente  $QA$  coincide con el coeficiente clásico  $Q(P) = \frac{B(P)}{I[\mathcal{N}_x]}$  que se usa para medir la calidad, en el sentido de adecuación a los datos, de la partición  $P_q$  [5, p. 155].

- Cuando el número de clases es menor o igual a  $s$ , los coeficientes  $QA$  y  $Q$  son diferentes. Si para una partición  $P_q$  ambos son pequeños, entonces  $P_q$  es una buena aproximación de una partición óptima  $P_q^*$ , a pesar de que  $P_q$  puede resumir mal los datos cuando hay una gran pérdida de inercia. Esta situación no es rara: para el ejemplo que será presentado más adelante, si se elige la métrica de Mahalanobis (inversa de la matriz de varianzas), tendremos  $QA(P_q) = 0,09$  y  $Q(P_q) = 0,13$ , con  $q = 2$ .

### 3. Orden sobre las particiones óptimas

El objetivo de esta sección es el de mostrar que la inercia interclases induce un orden, estricto hasta un cierto rango, sobre todo conjunto  $\{P_q^*/q = 1, \dots, n\}$  de particiones óptimas.

Sean  $q < n$  y  $P_q$  una partición de  $\Omega$  en  $q$  clases. Como existe al menos una clase  $P(k)$  en  $P_q$  que no sea reducida a un solo elemento, entonces se puede construir una nueva partición  $P_{q+1}^i$  en  $q + 1$  clases si se toma un elemento  $i$  de  $P(k)$  que forme una nueva clase y de la clase  $P(k)$  se elimina  $i$ , formando la clase  $P^i(k)$ , dejando igual las clases restantes.

Por el teorema de Huyghens (ver por ejemplo [9, p.59]), se tiene

$$B(P_{q+1}^i) = B(P_q) + \mu_i \|X_i - G_k\|_M^2 + \mu^i(k) \|G_k^i - G_k\|_M^2$$

donde  $G_k^i$  (resp.  $\mu^i(k)$ ) es el centro de gravedad (resp. peso) de la clase  $P^i(k)$ . Se deduce entonces que  $B(P_{q+1}^i) \geq B(P_q)$ .

Aplicando este procedimiento a las particiones óptimas, se obtiene entonces las sucesión de desigualdades:

$$0 = B(P_1^*) \leq B(P_2^*) \leq \dots \leq B(P_q^*) \leq \dots \leq B(P_n^*) = I[\mathcal{N}_x]. \quad (3)$$

Demostremos que este orden es estricto hasta un cierto rango, a partir del cual hay igualdad.

#### Propiedad 2

Una partición  $P_m$  de  $\Omega$  en  $m$  clases ( $m < n$ ) es tal que  $B(P_m) = I[\mathcal{N}_x]$  si y sólo si todos los puntos de cada clase coinciden con el centro de gravedad de su clase.

Además, se tiene  $m > s$ .

DEMOSTRACIÓN:

Sea  $B(P_m) = I[\mathcal{N}_x]$ . Si existe al menos un punto  $X_i$  que no coincida con el centro de gravedad de su clase, entonces por construcción de la partición  $P_{m+1}^i$  se tiene

$$I[\mathcal{N}_x] = B(P_m) < B(P_{m+1}^i) \leq I[\mathcal{N}_x]$$

lo cual es imposible.

Evidentemente, si todos los puntos coinciden con el centro de gravedad de su clase, por definición de la inercia se tiene  $B(P_m) = I[\mathcal{N}_x]$ .

Veamos ahora que el entero  $m$  es estrictamente mayor que  $s$ . En efecto, si esto no fuera cierto ( $m \leq s$ ), entonces  $\dim F_x \leq s - 1$  pues hemos supuesto que las variables son centradas, lo cual sería una contradicción con  $\dim F_x = \text{rang} X = s$ . ■

De la propiedad anterior y de la serie de desigualdades 3, demostramos la propiedad siguiente:

**Propiedad 3**

Para todo par de métricas  $M$  y  $D$ , existe un entero  $m$  tal que

$$0 = B(P_1^*) < B(P_2^*) < \dots < B(P_s^*) < \dots < B(P_m^*) = \dots = B(P_n^*) = I[\mathcal{N}_x]$$

DEMOSTRACIÓN:

Sea  $m$  el menor entero tal que existe una partición  $P_m$  de  $\Omega$  en  $m$  clases que cumple  $B(P_m) = I[\mathcal{N}_x]$  (si no hay puntos coincidentes entonces  $m = n$ ).

Por lo tanto  $B(P_m) = B(P_m^*)$  y se tiene entonces

$$B(P_m^*) = \dots = B(P_n^*) = I[\mathcal{N}_x].$$

Sea  $q < m$ . Sabemos que  $B(P_q^*) \leq B(P_{q+1}^*)$ .

Como  $q < n$ , siguiendo el procedimiento descrito al comienzo de esta sección, se puede construir, a partir de una partición óptima  $P_q^*$ , una partición  $P_{q+1}^i$  en  $q + 1$  clases aislando un elemento  $i$  de una clase no unitaria.

Se tiene entonces que  $B(P_{q+1}^i) > B(P_q^*)$ .

En efecto, si se diera el caso en que  $B(P_{q+1}^i) = B(P_q^*)$ , denotando por  $P_q^*(k)$  la clase de  $i$  en  $P_q^*$ , entonces se tendría que  $\forall j \in P_q^*(k) \quad \|X_j - G_k\|_M^2 = 0$ , donde  $G_k$  es el centro de gravedad de  $P_q^*(k)$ .

Todos los  $X_j$  coincidirían entonces con  $G_k$ .

Según la propiedad 2, se tendría entonces  $B(P_q^*) = I[\mathcal{N}_x]$  y por lo tanto  $q \geq m$ , lo cual es imposible por la forma como se escogió  $q$ . ■

De las propiedades anteriores, deducimos la consecuencia siguiente:

**Consecuencia:**

Las sucesiones  $(h_1(q))$  y  $(h_2(q))$  definidas por

$$\bullet h_1(1) = O \text{ y } h_1(q) = \sum_{j=1}^{s_q} \lambda_j^x \text{ para } q > 1;$$

$$\bullet h_2(q) = B(P_q^*)$$

son crecientes y para todo  $q$  se tiene  $h_1(q) \geq h_2(q)$ .

El  $q$ -ésimo término de  $h_2$  es la inercia explicada por el  $s_q$ -ésimo subespacio principal del ACP del triplete  $(\mathcal{N}_x, M, D)$ .

Para tener una idea de la forma en que crecen estas sucesiones y de la diferencia entre sus elementos, podemos representar sobre un mismo gráfico las funciones crecientes  $f$  y  $g$  de interpolación de los conjuntos de puntos  $\{(q, h_1(q)) / q = 1, \dots, n\}$  y  $\{(q, h_2(q)) / q = 1, \dots, n\}$  respectivamente (ver figura 1 más adelante).

**Nota:** en el caso en que se desee comparar las inercias explicadas por una partición y por un subespacio principal, las definiciones de las sucesiones anteriores permiten escoger la dimensión apropiada de este subespacio.



### Ejemplo de ilustración

Uno de los algoritmos usados para aproximar las particiones óptimas es el método de particiones principales [10, 16], que implementa un algoritmo de transferencias. Para tener una idea de los valores que toma el coeficiente  $QA$  sobre las aproximaciones dadas por este algoritmo, hemos considerado varias tablas de datos reales y simulados. En todos los ejemplos tratados, estos valores son en promedio pequeños (menores o iguales a 0.2). Estos resultados prueban la eficacia de este algoritmo sobre los datos utilizados.

Presentamos aquí los resultados obtenidos sobre un ejemplo. La tabla de datos es la usada por W. Castillo [4, p. 69] que contiene 29 individuos (cantones) descritos por 7 variables. Por simple consulta de la tabla, se puede notar que no hay puntos confundidos ( $m = n$ ). Todo conjunto de particiones óptimas está por lo tanto estrictamente ordenado por la inercia explicada. Es más, como las variables centradas son linealmente independientes, la matriz de varianzas-covarianzas  $V$  y su inversa  $V^{-1}$  (matriz de Mahalanobis) son definidas positivas.

Escogiendo  $M = V$  y pesos iguales, y usando el método de particiones principales mediante transferencias, hemos aproximado 9 de las primeras particiones óptimas. Todas las particiones son inicializadas al azar.

Las aproximaciones  $\{\tilde{P}_q/q = 1, \dots, 9\}$  de las 9 primeras particiones óptimas son tales que:

$$0 = B(\tilde{P}_1) < B(\tilde{P}_2) < \dots < B(\tilde{P}_9).$$

Los valores del coeficiente  $QA$ , para todas las particiones anteriores, son menores a 0,25.

Puesto que no se conocen los valores de la sucesión  $(h_2(q))$ , hemos aproximado, mediante una función de interpolación  $g$ , la sucesión creciente de puntos

$$\left\{ \left(1, B(\tilde{P}_1)\right), \dots, \left(9, B(\tilde{P}_9)\right), (29, I[\mathcal{N}_x]), \right\}.$$

Esta función de aproximación y la función  $f$  de interpolación del conjunto

$$\{(1, h_1(1)), \dots, (9, h_1(9)), (29, h_1(29))\}$$

son representadas en la figura 1. Este gráfico muestra que las inercias explicadas por las particiones óptimas en  $q$  clases son muy cercanas a las inercias explicadas por los subespacios principales de dimensiones  $s_q$ .

## 4. Problema de la escogencia del número de clases

La importancia y la dificultad del problema de la escogencia del número de clases ha sido señalada por diversos autores [5, 7, 8, 11, 14, etc.] que se han interesado en este problema. Everitt [11] estudia y compara diferentes métodos propuestos para tratar de resolver este problema.

La definición de la sucesión  $(h_2(q))$  muestra que toda partición óptima es mejor que las particiones en un número menor de clases, en el sentido del criterio de la inercia interclases. Por lo tanto, este problema no es matemático sino estadístico. Su dificultad radica esencialmente en dos objetivos contrarios que se persiguen en Clasificación Automática:

*la búsqueda de una partición cuyo número de clases sea lo más pequeño posible y cuya inercia interclases sea lo más grande posible.*

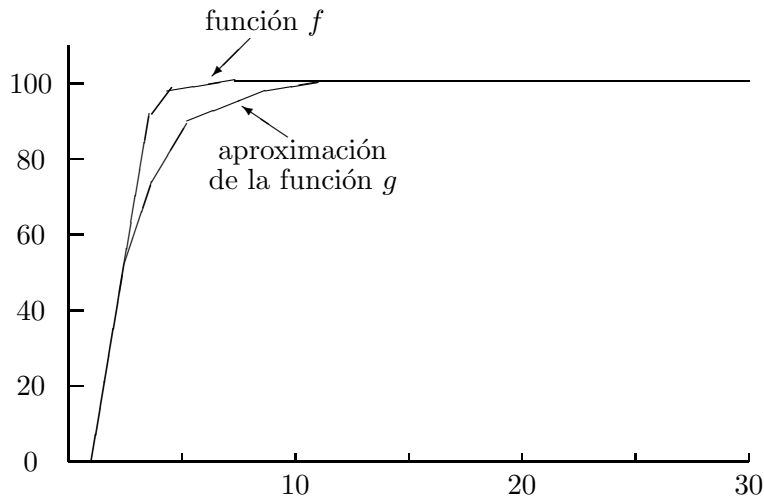


Figura 1: funciones de interpolación

De las definiciones de las sucesiones  $(h_1(q))$  y  $(h_2(q))$ , deducimos que una partición óptima que verifique el segundo objetivo debe ser tal que:

$$\frac{B(P_q)}{B(P_{(q+1)})} \simeq \frac{h_1(q)}{h_1(q+1)} \simeq 1.$$

El número de clases  $q^*$  mejor adaptado a los datos tratados será por lo tanto el menor valor tal que el cociente  $\frac{h_1(q^*)}{h_1(q^*+1)}$  es cercano a 1.

Este valor corresponde al punto a partir del cual la función  $f$  es prácticamente constante. Puede ser determinado ya sea al examinar el gráfico de  $f$ , o bien al examinar el conjunto de cocientes  $\{h_1(q)/h_1(q+1)/q = 1, \dots, s+1\}$ .

El número  $q^*$  corresponde también a la menor dimensión  $s_{q^*}$  de los subespacios principales más significativos. Esta proposición heurística está acorde con la escogencia clásica del número de clases, que Lerman [17, pág. 331] resume con la frase:

*“en general,  $q$  es sensiblemente más grande que el número de factores interpretables”.*

Para un estudio más detallado de los datos, sería útil examinar las aproximaciones de las particiones óptimas cuyo número de clases es menor que  $q^*$  o igual a  $q^* + 1$ .

## Aplicación

Para el ejemplo de ilustración presentado más arriba, la sucesión de cocientes  $(h_1(q)/h_1(q+1)/q = 1, \dots, s+1)$  es  $(0,9249, 0,9903, 0,9940, \simeq 1, \simeq 1, \simeq 1, 1)$ . Los elementos de esta sucesión son casi todos iguales a 1. Observamos que a partir del índice 3, sus variaciones son muy pequeñas. El número de clases adaptado a los datos y sugerido por nuestra proposición, es entonces 4.

En la figura 2 se representan los individuos en el primer plano principal, que explica 99,44% de la inercia total de la nube de puntos  $\mathcal{N}_x$ . Hemos representado mediante un trazo continuo las cuatro clases de  $\tilde{P}_4$ . Dos de estas clases se subdividen a su vez (con línea punteada) para completar las seis clases de  $\tilde{P}_6$ , llamadas  $C_1, \dots, C_6$ .

Estas clases son:

- $C_1 = \{\text{Santa Cruz, Cañas, Puriscal, Limón, Santa Bárbara, San Rafael, Aserrí,}$   
 $\quad \quad \quad \text{Ñicoya, Tarrazú, Atenas, Puntarenas, Esparza, Mora, Liberia, San Mateo}\}$
- $C_2 = \{\text{Barva, La Unión}\}$
- $C_3 = \{\text{Grecia}\}$
- $C_4 = \{\text{Paraíso, Cartago, Escazú, San Ramón, Alajuela}\}$
- $C_5 = \{\text{Santo Domingo, Palmares, Heredia, Naranjo, Desamparados}\}$
- $C_6 = \{\text{San José}\}$

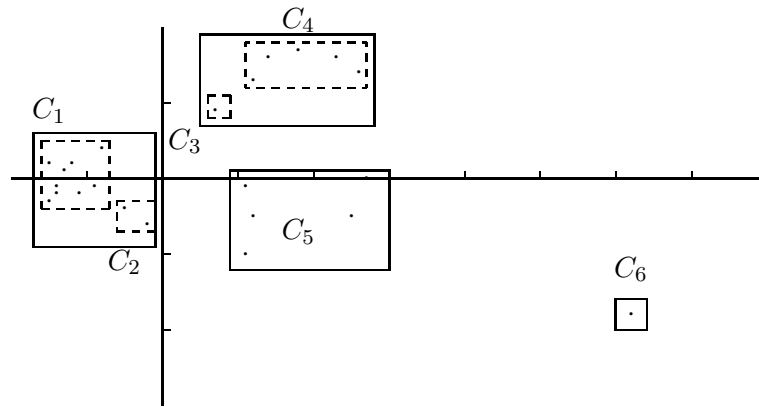


Figura 2: Particiones representadas en el primer plano principal

Hay que notar que sobre el primer plano principal las clases están bien separadas, lo que permite una buena interpretación de los datos. Las particiones  $\tilde{P}_1, \dots, \tilde{P}_6$  están encajadas. Este hecho lo hemos representado en la figura 3, que muestra la jerarquía que forman las particiones a partir de los conjuntos  $C_1, \dots, C_6$ . En esta figura, hemos presentado la inercia interclases de cada partición. Así, las seis primeras particiones obtenidas son:

$$\begin{aligned} \tilde{P}_1 &= \Omega \\ \tilde{P}_2 &= (C_1, C_2, C_3, C_4), (C_5, C_6) \\ \tilde{P}_3 &= (C_1, C_2, C_3, C_4), (C_5), (C_6) \\ \tilde{P}_4 &= (C_1, C_2), (C_3, C_4), (C_5), (C_6) \\ \tilde{P}_5 &= (C_1, C_2), (C_3), (C_4), (C_5), (C_6) \\ \tilde{P}_6 &= (C_1), (C_2), (C_3), (C_4), (C_5), (C_6). \end{aligned}$$

### Observación

El problema del número de clases no se plantea en el Análisis Factorial Relacional, pues se muestra que la partición trivial, con  $q = n$ , no es necesariamente óptima [19]. Una de las ventajas de este método es que el número de clases de la partición buscada es uno de los resultados del algoritmo.

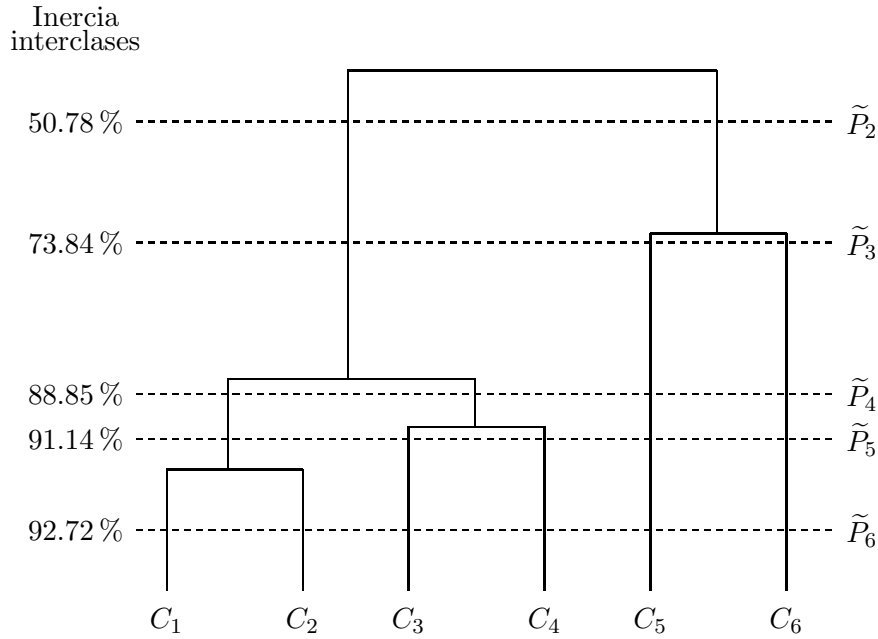


Figura 3: Jerarquía formada por las primeras aproximaciones de las particiones óptimas

## 5. Conclusión

Los pequeños valores del coeficiente  $QA$  muestran que, en más de 300 ejemplos de datos reales o simulados, la inercia explicada por el  $s_q$ -ésimo subespacio principal es cercana a la inercia interclases de una partición óptima  $P_q^*$ . ¿Son esta observación y los buenos resultados obtenidos hasta el momento, generales? es decir, ¿son independientes de los datos tratados y de las métricas escogidas, o bien específicos a las escogencias hechas?

Las respuestas a estas preguntas, que tienen que ver con la relación entre los resultados del ACP y los de la clasificación, nos parecen difíciles. La dificultad radica en que el primer problema posee una solución formal (analítica) y el segundo es un problema combinatorio. Para responder, aunque sea parcialmente, a estas interrogantes, las investigaciones presentadas aquí deben ser profundizadas y completadas. En particular, habría que considerar otros conjuntos de datos, otras métricas y escoger otros algoritmos de clasificación.

## Agradecimientos

*La dirección y las sugerencias del Profesor Y. Schektman, han sido de un gran valor científico en nuestras investigaciones, por lo que queremos manifestar aquí nuestro reconocimiento y agradecimiento.*

## Referencias

- [1] Bédécarrax, C.; Huot, C. (1991) *Développement d'indicateurs pour l'interprétation des résultats d'une analyse factorielle-relationnelle*. Etude MAP-005, Décembre 1991, Centre Européen de Mathématiques Appliquées, Compagnie IBM-France.
- [2] Benzécri, J.-P. y colaboradores (1973) *Analyse des Données*, tomo II. Dunod, París.
- [3] Caillez, F.; Pagès, J.-P. (1976) *Introduction à l'Analyse des Données*. SMASH, París.
- [4] Castillo, W. (1991) *Descripción de algunos métodos de clasificación automática y aplicación a un problema de producción distribuida por cantón*. Ciencias Matemáticas, Vol. II, N<sup>o</sup> 1, pp.67-78.
- [5] Celeux, G.; Diday, E.; Govaert, G.; Lechevallier, Y.; Ralambondrainy, H. (1989) *Classification Automatique des Données*. Dunod-Informatique, París.
- [6] Chandon, J.L.; Pinson, S. (1981) *Analyses Typologiques, Théories et Applications*. Masson, París.
- [7] Cormack, M. (1971) *A review of classification*. Journal of the Royal Statistical Society, serie A, 134, N<sup>o</sup> 3, pp. 321-367
- [8] Diday, E. y colaboradores (1980) *Optimisation en Classification Automatique*. INRIA, Rocquencourt.
- [9] Diday, E.; Lemaire, J.; Pouget, J.; Testu, F. (1985) *Eléments d'Analyse de Données*. Dunod, París.
- [10] Espinoza, J.L.; Trejos, J. (1989) *Clasificación por particiones*. Revista Ciencia y Tecnología, Vol. XIII, Nos. 1-2, pp.129-154.
- [11] Everitt, B.S. (1979) *Unresolved problems in cluster analysis*. Biometrics, 35, pp. 169-181.
- [12] Friedman, H.P.; Rubin, J. (1967) *On some invariant criteria for grouping data*. Journal of the American Statistical Association, 1967, 62, pp. 1159-1178.
- [13] Gondran, M. (1976) *Valeurs propres et vecteurs propres en classification hiérarchique*. RAIRO, Recherche Opérationnelle, serie R: Informatique Théorique, Vol. 10, n<sup>o</sup> 3, pp. 39-46.
- [14] Gower, J.C. (1973) *Classification problems*. Bulletin de l'Institut International de Statistique, Actes de la 39ème session, Viena.
- [15] Gower, J.C. (1974) *Maximal predictive classification*. Biometrics, Vol. 30, pp. 643-654.
- [16] Ibrahim, A.; Schektman, Y. (1986) *Principal cluster analysis*. En: Classification as a Tool for Research, W. Gaul and M. Schader (eds.), Elsevier Sc.Publ, North-Holland, pp. 217-233.
- [17] Lerman, I.C. (1979) *Les présentations factorielles de la classification*. RAIRO, Vol. 13 N<sup>o</sup>2, p.107-128 y N<sup>o</sup>3, p.227-251.
- [18] Lerman, I.C. (1981) *Classification et Analyse Ordinale des Données*. Dunod, París.
- [19] Marchotorchino, F. (1991) *L'analyse factorielle-relationnelle, I y II*. Etude MAP-003, Centre Européen de Mathématiques Appliquées, Compagnie IBM-France.
- [20] Marchotorchino, F.; Bédécarrax, C. (1992) *Le critère de différence de profils*. Distancia'92, Rennes.
- [21] Mirkin, B.G. (1987) *Additive clustering and qualitative factor analysis. Methods for similarity matrices*. Journal of Classification, N<sup>o</sup> 1, Springer-Verlag, New-York, pp. 3-27

- [22] Nashed, Z., editor (1976) *Generalized Inverses and Applications*. Academic Press, London.
- [23] Rao, C.R (1964) *The use and interpretation of principal component analysis in applied research*. Sankhya, serie A, 1964, 26, pp. 329-358.
- [24] Schektman, Y. (1978) *Contribution à la mesure en facteurs dans les sciences expérimentales et à la mise en œuvre automatique dans les calculs statistiques*. Tesis de Estado, Toulouse.