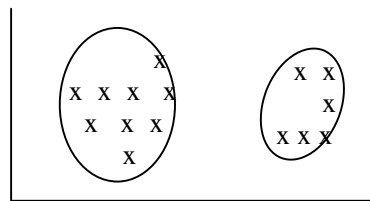


Introducción

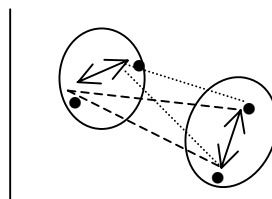
Nuestros datos, desde el punto de vista geométrico, configuran una nube de puntos en un espacio (recordemos la dualidad entre el espacio de los casos y el de las variables por la que la que esta configuración podría ser la nube de los casos en el espacio de las variables o la nube de las variables en el espacio de los casos). Pretendemos clasificar los puntos de esta nube en grupos; y nuestra idea intuitiva de grupo o conglomerado nos llevaría a acordar que los elementos de cada grupo deben estar relativamente juntos y presentar poca distancia entre sí como indicativo de tener comportamientos (coordenadas) similares, mientras que los que pertenezcan a grupos distintos se deben encontrar relativamente más distantes o separados como indicativo de tener comportamientos (coordenadas) menos parecidos.



El Concepto de Conglomerado

Se han dado algunas definiciones para intentar formalizar esa idea intuitiva de “grupo” que se conoce habitualmente como *Conglomerado* en terminología estadística española, o como *Cluster* si asumimos la terminología anglosajona; cuando no nos referimos a ellos coloquial y simplemente como *Grupo*.

Así y a modo de ejemplos, Kendall y Buckland definen los conglomerados como *conjuntos de elementos contiguos de una población estadística*. La “contigüidad” es en esta definición la que expresa la similitud entre los individuos observados. Por su parte, Gengrelli en 1963 define el conglomerado como un *grupo de elementos tales que la distancia entre cada dos puntos es menor que la distancia de cualquiera de ellos a algún elemento de otro conglomerado*. Dicho de otra manera, son elementos que están muy cercanos entre sí y tales que las distancias entre ellos son menores que cualquier distancia a elementos de otros conglomerados como se muestra en el siguiente gráfico:



Wallace y Boulton, en 1968, nos definen un conglomerado como un *grupo de elementos que pueden tratarse como equivalentes en algún sentido*. Esta definición, más finalista, pone su énfasis en que, una vez que tenemos los elementos clasificados en grupos relativamente homogéneos de elementos parecidos entre sí, entonces los elementos de cada

2 ANÁLISIS MULTIVARIANTE DE DATOS

grupo pueden ser considerados como equivalentes al presentar comportamientos similares, mientras que elementos de grupos distintos serán considerados siempre distintos (de diferentes comportamiento).

Son estos grupos, conglomerados o clusters los que se pretenden descubrir y determinar mediante el Análisis de Conglomerados.

El Problema General

El problema general del Análisis de Conglomerados lo podemos expresar resumidamente de la siguiente forma:

Dado un conjunto de n objetos, elementos o individuos $I = \{I_1, I_2, \dots, I_n\}$ (que recordemos pueden ser casos o variables dependiendo del espacio considerado) sobre los que se han medido p características X_1, X_2, \dots, X_p , (si los individuos son casos estas características serán variables y si los individuos son variables entonces las características serán sus observaciones sobre los casos) se quiere formar k conglomerados G_1, G_2, \dots, G_k , (k no necesariamente determinado a priori), de forma que, en algún sentido, los individuos emplazados en un mismo conglomerado $G_i, i=1,2,\dots,k$ sean más homogéneos entre sí que con los individuos procedentes de otros grupos.

El proceso mediante el cual se asigna un individuo a un determinado conglomerado en base a las características de éste y aquél, se denomina Identificación

Elementos del Análisis de Conglomerados

La aplicación de las técnicas del Análisis de Conglomerados requiere la consideración y cumplimiento metódico de una serie de etapas:

Elección de los objetos a analizar, que queremos clasificar y cuya elección dependerá del estudio que estemos realizando. Aunque estos individuos pueden ser casos o variables, en el desarrollo general de este tema supondremos que vamos a clasificar casos en el espacio de las variables; y haremos notar lo necesario cuando consideremos la clasificación de variables.

Elección de las características observables que interesan para el estudio. Una vez elegidos los objetos, notaremos en general que sobre cada uno de esos casos se observan un amplio número de variables. Puede que para el estudio de similaridad no sean necesarias todas ellas sino solamente aquéllas que recogen la información necesaria para evaluar su proximidad de acuerdo con los objetivos planteados.

Adopción del concepto de “similitud” entre objetos adaptado al problema de acuerdo con los planteamientos teóricos y objetivos del estudio, y *determinación de la Medida de la similitud/disimilaridad entre objetos* buscando su consistencia con el contenido conceptual de la definición de proximidad adoptado en el problema, bien mediante su definición directa derivada de los planteamientos teóricos realizados, o bien por elección de alguna formulación de entre todas las que han sido profusamente estudiadas en el tema dedicado a las medidas de proximidad.

Preparación de datos y Homogeneización de variables Una vez decididas qué características son las que nos interesan para nuestro estudio deberemos preparar los datos y homogeneizar las variables para que todas se midan en escalas compatibles con la medida de similitud o disimilaridad que expresará el grado de proximidad entre los individuos. Generalmente, se exigirá que todas las variables vengan expresadas en un mismo tipo de escala de medida.

Adopción del concepto de “conglomerado” adaptado al problema y elección de la Técnica para la formación de conglomerados. Como veremos en este capítulo, existe una amplia gama de técnicas que producen conglomerados con distintas lógicas y formas, por lo que deberemos tener la precaución de elegir la técnica adecuada, siempre en función de la aproximación racional que hagamos al concepto de conglomerado adoptado en el problema.

Es en este punto anterior donde encontrará aplicación lo que veremos en este capítulo en el que se revisarán las principales técnicas de formación de conglomerados que, basadas en la disponibilidad de alguna medida de proximidad (similaridad o disimilaridad) van a proporcionar la clasificación de los datos en grupos o conglomerados.

Interpretación de resultados y revisión de planteamientos. Por supuesto que como en todo proceso científico, al final, una vez que tengamos la distribución de grupos obtenida, habrá que interpretar los resultados y contrastarlos de alguna manera con las evidencias aceptadas, para valorar la necesidad, o no, de revisar la adecuación del planteamiento adoptado para el problema, la idoneidad de la medida de proximidad y de la técnica de clasificación elegidas, etc., con el objeto de que, en caso necesario, podamos intentar mejorar el proceso.

Técnicas de Clasificación Jerárquicas

El proceso de agrupación de objetos, tal y como podemos entender de una forma intuitiva, pivota sobre la tolerancia o exigencia que deseemos sobre la proximidad de los objetos dentro de los grupos y la proximidad entre los distintos grupos o conglomerados. Cuanto más exigentes seamos al respecto (cuanto más homogéneos deban ser los objetos clasificados en un mismo conglomerado), deberemos considerar un mayor número de grupos para distribuir la diversidad observada, obteniéndose mucha homogeneidad interna en dichos conglomerados. Por el contrario, cuanto más tolerantes seamos respecto a la proximidad (cuanto más permisivos seamos con las diferencias para que los objetos puedan ser clasificados en un mismo conglomerado), menor número de grupos serán necesarios para recoger toda la diversidad de comportamientos, pero más heterogéneos serán los conglomerados en sus interiores.

Por ello, según exijamos mayor (o menor) homogeneidad interna a los conglomerados en que clasificaremos los objetos estudiados, obtendremos diferentes clasificaciones con un mayor (o menor) nivel de detalle diferenciador de cada conglomerado resultante y un mayor (o menor) número de éstos.

Así, a partir de cualquier estructura de conglomerados obtenida a partir de una cierta clasificación, podríamos intentar extraer otra más detallada, con un mayor número de grupos más específicos, sin más que exigir mayor homogeneidad interna en los grupos

4 ANÁLISIS MULTIVARIANTE DE DATOS

observando, por ejemplo, la respuesta de cierta característica que los pueda diferenciar.

Cuando los nuevos grupos así derivados quedan perfectamente encajados en el conglomerado de procedencia, de forma que todos los objetos de los nuevos conglomerados provienen sin excepción del conglomerado anterior, el procedimiento de clasificación recibe el nombre de *Jerárquico*.

Matemáticamente, dado un espacio E formado por una cierta cantidad de individuos a clasificar, $E = \{1, 2, \dots, n\}$, entonces se dice que un conjunto de las partes de E , $H \in \wp(E)$ es una Jerarquía, si y solo si verifica los dos siguientes axiomas:

$$\text{Axioma de la intersección: } \forall h, h' \in H \Rightarrow h \cap h' \in \{h, h', \phi\}$$

$$\text{Axioma de la unión: } \forall h \in H \Rightarrow \bigcup \{h' \mid h' \in H, h' \subset h\} \in \{h, \phi\}$$

Si $E = \{1, 2, \dots, n\} \in H$ y $\{i\} \in H \quad \forall i \in E$, entonces H recibe el nombre de *Jerarquía Total*.

Si además, existe una función $d: H \rightarrow \mathcal{R}^+$, tal que verifique que:

$$\text{a) } d(\{i\}) = 0 \quad \forall i \in E$$

$$\text{b) si } h \subset h' \Rightarrow d(h) < d(h')$$

entonces a H se le conoce con el nombre de *Jerarquía Indexada* y a d se le llama *índice de la jerarquía*.

Un ejemplo típico de una clasificación jerárquica podría ser la siguiente Clasificación simplificada de los Animales:



en la que todos los vertebrados o invertebrados, sin excepción, son animales aunque se diferencian en la posesión o no de algún tipo de esqueleto; o en la que todos los mamíferos, aves, reptiles, anfibios y peces son vertebrados aunque se diferencian en si son de sangre fría o caliente y en el tipo de recubrimiento de la piel; etc.

Los *algoritmos de clasificación jerárquica* van a proveer clasificaciones jerárquicas organizadas en *niveles* obtenidos de forma secuencial, de forma que a mayor profundidad del nivel, mayor especificidad (homogeneidad interna más estrecha) de los objetos

clasificados y mayor número de grupos diferentes; y viceversa.

Cuando los distintos niveles se obtiene tal y como hemos observado en el ejemplo anterior, “dividiendo” jerárquicamente cada cluster o conglomerado, en cada etapa, en función de que se dé o no una o varias características en los objetos (por ejemplo, “existencia de esqueleto” o “tipo de sangre y recubrimiento de la piel”) los algoritmos se denominan *algoritmos divisivos de clasificación jerárquica*. Si para estas divisiones de cada grupo de un nivel anterior sólo se emplea una característica, el algoritmo recibe el calificativo de *Divisivo Monotético*. Por el contrario, si se emplean varias características simultáneamente, el algoritmo se recibe el calificativo de *Divisivo Politético*.

Los algoritmos divisivos parten, pues de disponer a todos los objetos agrupados en un solo conjunto y, mediante divisiones sucesivas, consiguen desagregar a los elementos tanto como sea deseado, pudiendo llegar a un último nivel en el que habría tantos conglomerados como objetos, donde los conglomerados estarían formados por un único objeto y donde las diferencias entre conglomerados serían pues las diferencias individuales.

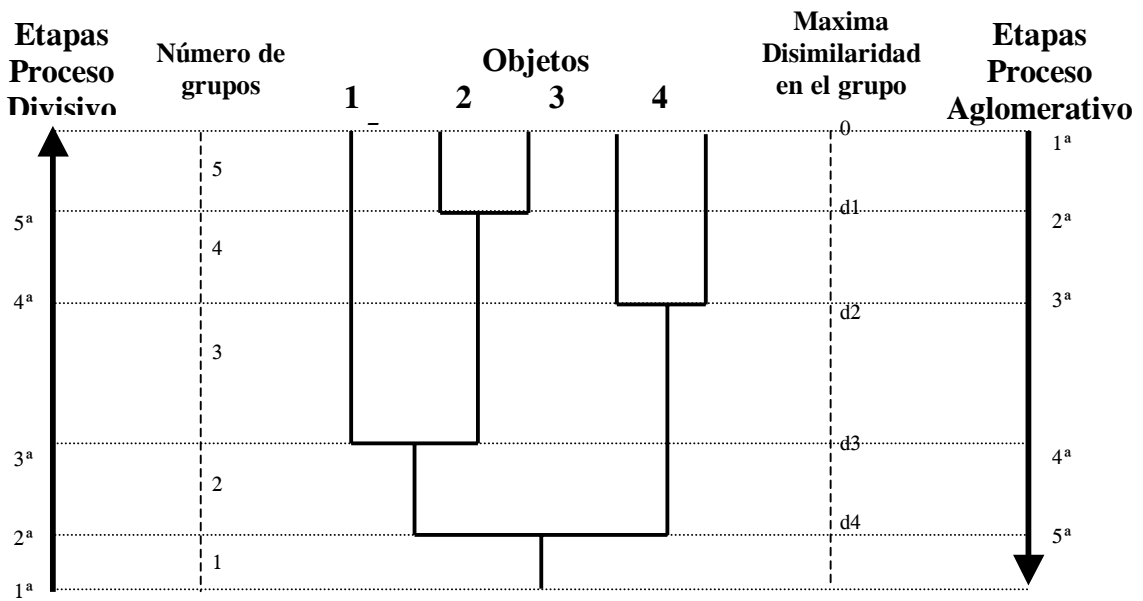
Cuando los distintos niveles se obtiene justo en sentido contrario, es decir, “agregando” jerárquicamente varios clusters o conglomerados (generalmente dos), en cada etapa, en función de ser los más próximos o similares, los algoritmos se denominan *algoritmos agregativos de clasificación jerárquica*. Así se va obteniendo, en cada etapa, una nueva clasificación con un menor número de grupos, pero con una mayor diversidad o heterogeneidad interna que en la etapa anterior.

Los algoritmos aglomerativos parten, pues de disponer a todos los objetos considerados individualmente y, mediante uniones sucesivas, consiguen llegar finalmente a agrupar a todos los objetos en un único conglomerado, donde serían indistinguibles al haber tolerado fusionar en un mismo cluster incluso a los objetos que presentaron la máxima disparidad. O sea, que en la primera etapa, habrá n grupos; en la segunda etapa, $(n-1)$ grupos, y así sucesivamente hasta llegar al grupo único. En general en las etapas intermedias vamos a encontrar clasificaciones con k grupos sea cual sea el número k comprendido entre 1 y el número total de objetos, n .

Por ejemplo, si partimos de 5 elementos $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$ y $\{5\}$ y decidimos unir el elemento $\{2\}$ con el $\{3\}$, en la segunda etapa tendríamos un grupo $\{2, 3\}$ y otros grupos con cada uno de los elementos restantes: $\{1\}$, $\{4\}$ y $\{5\}$. En total tendríamos $4=5-1$ grupos, pues lo único que hemos hecho ha sido fusionar dos de los elementos de la etapa anterior en uno nuevo. En una clasificación jerárquica, una vez que los elementos $\{2\}$ y $\{3\}$ se unen en un mismo grupo, irán juntos en todas las etapas posteriores, y no se volverán a separar.

Podemos representar esquemáticamente las etapas de estos procesos, tanto agregativos como divisivos, a través del llamado *dendrograma*, como sigue:

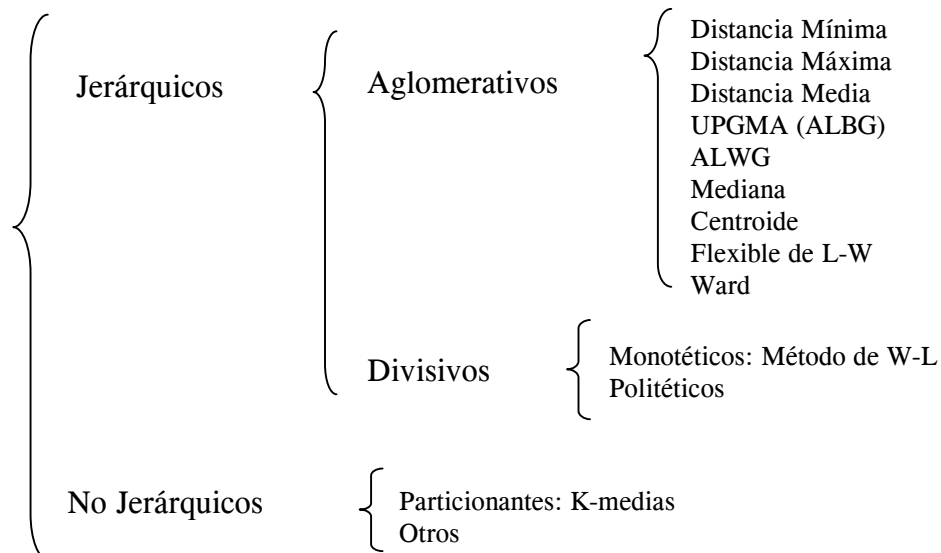
6 ANÁLISIS MULTIVARIANTE DE DATOS



La clasificación que generalmente nos interesará será probablemente una de las intermedias. El decidir la etapa en que nos pararemos en este proceso dependerá de las exigencias que impongamos sobre la afinidad entre los individuos dentro de cada grupo, así como sobre la diferencia entre individuos de grupos distintos.

El problema fundamental de las técnicas jerárquicas será su gran exigencia de computación, lo que, cuando el numero de variables y casos es importante, debemos recurrir a otras técnicas de clasificación, generalmente basadas en asignación óptima (mediante optimización matemática), para los que se suele fijar de antemano el número de conglomerados a identificar y para los que puede ocurrir que individuos que estén juntos en un mismo grupo en una determinada etapa de identificación, se reasignen en grupos diferentes en otra etapa posterior. Estos métodos se conocen como *métodos de optimización no jerárquicos*.

A continuación presentamos las principales técnicas de clasificación que vamos a ver a continuación, clasificadas según su tipología:



Algoritmo fundamental para la obtención de Clasificaciones Jerárquicas Indexadas

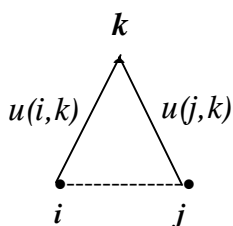
Este algoritmo se aplica generalmente en técnicas agregativas de clasificación jerárquicas.

Iteración 0 (situación inicial): El proceso comienza considerando la partición C_0 : $\{1\}, \{2\}, \dots, \{n\}$ del conjunto de objetos a clasificar, formada por los n objetos a clasificar considerados individualmente. Además, debemos disponer de una ultramétrica u definida sobre C_0 que exprese la proximidad existente entre los n objetos a clasificar. Es decir, en principio, cada objeto se considera como un grupo aislado y la ultramétrica u expresa la proximidad existente entre cada par de ellos.

Entonces, estaremos en condiciones de iniciar el proceso iterativo de clasificación como se refleja a continuación, de forma genérica, para una iteración cualquiera r -ésima, pudiendo r tomar un valor entero entre 1 y n .

Iteración r -ésima, paso 0: Tras las $r-1$ iteraciones anteriores ($r=1, 2, \dots, n$), dispondremos de una partición del conjunto de objetos a clasificar C_{r-1} : $h_1^{(r-1)}, \dots, h_{n-(r-1)}^{(r-1)}$, y de una ultramétrica u sobre C_{r-1} . (Obsérvese que esto es evidente para $r=1$ si tenemos en cuenta lo exigido como situación inicial en la iteración 0. En las demás iteraciones veremos que es una consecuencia del proceso que sigue este algoritmo fundamental).

Recordemos que si tenemos una ultramétrica definida sobre un conjunto de objetos, entonces, si representamos gráficamente 3 cualesquiera de ellos de forma que la distancia entre cada dos de éstos sea la correspondiente proximidad que expresa la ultramétrica, entonces los 3 objetos siempre se encontrarán formando un triángulo isósceles cuya base corresponde a los dos puntos más cercanos, como se indica esquemáticamente en el siguiente dibujo:



Iteración r -ésima, paso 1: Se determinan aquellos dos objetos que presentan una proximidad mínima y supongamos que son los de la pareja $(h_i^{(r-1)}, h_j^{(r-1)})$. Es decir,

$$u(h_i^{(r-1)}, h_j^{(r-1)}) = \min_{l \neq k} \{u(h_l^{(r-1)}, h_k^{(r-1)})\}$$

Lo que perseguimos será unir en esta iteración estos dos objetos para formar un nuevo grupo agregado $h_i^{(r-1)} \cup h_j^{(r-1)}$. Y como con cualquier otro formarían un triángulo isósceles para u , por ser u una ultramétrica, podremos definir la distancia del nuevo grupo a cualquier

8 ANÁLISIS MULTIVARIANTE DE DATOS

elemento $h_k^{(r-1)}$ como cualquiera de las $u(h_i^{(r-1)}, h_k^{(r-1)})$ ó $u(h_j^{(r-1)}, h_k^{(r-1)})$ ya que van a ser iguales.

Iteración r-ésima, paso 2: Así pues, se procede a agregar los dos objetos más próximos y se forma la nueva partición, a partir de la que teníamos:

$$C_r: h_1^{(r-1)}, \dots, h_i^{(r-1)} \cup h_j^{(r-1)}, \dots, h_{n-(r-1)}^{(r-1)} \equiv h_1^{(r)}, \dots, h_{n-r}^{(r)}$$

En cada iteración reduciremos el número de grupos en uno. Así, partimos de n grupos; al terminar la primera iteración tendremos n-1; al terminar la segunda, n-2; y al final, tras la iteración n-1, nos quedará un único grupo formado por todos los objetos.

Iteración r-ésima, paso 3: Construir una nueva ultramétrica \bar{u} sobre la nueva partición C_r de la siguiente forma:

$$\begin{cases} \bar{u}(h_k^{(r-1)}, h_i^{(r-1)} \cup h_j^{(r-1)}) = u(h_k^{(r-1)}, h_j^{(r-1)}) = u(h_k^{(r-1)}, h_i^{(r-1)}) \\ \bar{u}(h_k^{(r-1)}, h_l^{(r-1)}) = u(h_k^{(r-1)}, h_l^{(r-1)}), \quad k, l \neq i, j \end{cases}$$

Por ser u una ultramétrica se puede demostrar que esta nueva medida \bar{u} vuelve a ser también otra ultramétrica y, por tanto, el proceso puede reiterarse rigurosamente, comenzando con el paso 0 de una nueva iteración.

La demostración es sencilla y bastará con ver que la nueva medida cumple con la propiedad ultramétrica. Si consideramos 3 grupos u objetos distintos de los fusionados en este paso, el cumplimiento de esta propiedad es evidente por ser la nueva ultramétrica para cualesquiera dos de esos 3 puntos exactamente lo que valía la antigua ultramétrica. Sólo debemos centrarnos en el caso en que uno de los 3 objetos puntos es el nuevo conjunto fusionado; en cuyo caso, las dos situaciones no tan evidentes serían:

$$\begin{aligned} \bar{u}(h_k^{(r-1)}, h_l^{(r-1)}) &= u(h_k^{(r-1)}, h_l^{(r-1)}) \leq \max\{u(h_k^{(r-1)}, h_i^{(r-1)}), u(h_l^{(r-1)}, h_i^{(r-1)})\} = \\ &= \max\{\bar{u}(h_k^{(r-1)}, h_i^{(r-1)} \cup h_j^{(r-1)}), \bar{u}(h_l^{(r-1)}, h_i^{(r-1)} \cup h_j^{(r-1)})\} \end{aligned}$$

y

$$\begin{aligned} \bar{u}(h_k^{(r-1)}, h_i^{(r-1)} \cup h_j^{(r-1)}) &= u(h_k^{(r-1)}, h_i^{(r-1)}) \leq \max\{u(h_k^{(r-1)}, h_l^{(r-1)}), u(h_i^{(r-1)}, h_l^{(r-1)})\} = \\ &= \max\{\bar{u}(h_k^{(r-1)}, h_l^{(r-1)}), \bar{u}(h_i^{(r-1)} \cup h_j^{(r-1)}, h_l^{(r-1)})\} \end{aligned}$$

como se pretendía demostrar.

Este procedimiento produce clasificaciones jerárquicas ya que, una vez que se han unido unos objetos, siempre permanecen unidos en los grupos que se vayan creando posteriormente

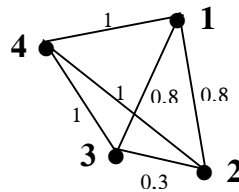
Ejemplo:

Partamos de cinco objetos entre los que se ha observado la medida de disimilaridad expresada en la siguiente tabla:

	1	2	3	4	5
1	0,0	0,8	0,8	1,0	1,0
2		0,0	0,3	1,0	1,0
3			0,0	1,0	1,0
4				0,0	0,5
5					0,0

Obsérvese que la disimilaridad de cualquier individuo consigo mismo es cero (ceros en toda la diagonal principal) y la disimilaridad entre objetos distintos será mayor o menor según indican los valores extra-diagonales.

En la práctica, disponer de ultramétricas no es frecuente; sin embargo, la expresada en la tabla anterior lo es. Así, por ejemplo, si representáramos los objetos 1, 2, 3 y 4 (obviamente sólo podremos representar hasta 4 puntos en un espacio de dimensión 3 que podamos visualizar intuitivamente) obtendríamos una representación como la siguiente:



Se puede observar que todos los triángulos formados por cada tres objetos son isósceles; lo que volvería a ocurrir si añadiéramos el objeto 5 en el espacio de una dimensión más. Luego, efectivamente, la matriz anterior expresa realmente una ultramétrica: cada tres puntos comparados formarían un triángulo isósceles.

La iteración 0 (o inicial) simplemente concreta la situación de partida, que se reduce a comenzar con la partición de los objetos en los 5 grupos individuales $\{1\}$ $\{2\}$ $\{3\}$ $\{4\}$ $\{5\}$ y sobre la que está definida la ultramétrica anterior.

Iteración 1ª:

Según nos dice el paso 1, buscaremos los 2 objetos más próximos, los más parecidos; que serán los individuos 2 y 3 con un nivel de disimilaridad de 0,3.

El paso 2 nos conduce a unirlos como nuevo grupo, por lo que obtendríamos la clasificación en 4 grupos $\{1\}$ $\{2,3\}$ $\{4\}$ $\{5\}$

Y con el paso 3 obtendríamos la ultramétrica inducida sobre él, que sería la expresada en la siguiente tabla.

	1	2;3	4	5
1	0,0	0,8	1,0	1,0
2;3		0,0	1,0	1,0
4			0,0	0,5
5				0,0

O sea, las disimilaridades entre elementos distintos de los que hemos unido siguen obviamente siendo las mismas (por ejemplo, $d(1,4)$ sigue siendo 1.0, $d(1,5)$ sigue siendo 1.0 y $d(4,5)$ seguirá siendo 1.0) y la disimilaridad entre cualquier otro objeto con el grupo que acabamos de formar por agregación de los objetos 2 y 3, sería la que presenta aquél con cualquiera de los dos objetos

fusionados ya que, al ser una ultramétrica, serán iguales (por ejemplo, la $d(1,2)$ y la $d(1,3)$ eran ambas 0.8 y por tanto $d(1, \{2,3\})$, será 0.8).

10 ANÁLISIS MULTIVARIANTE DE DATOS

Iteración 2ª:

Paso 1: Ahora, los 2 objetos más próximos serán los individuos 4 y 5 con un nivel de disimilaridad de 0,5.

Paso 2: La nueva clasificación en 3 grupos sería $\{1\} \{2,3\} \{4,5\}$

Paso 3: Y la nueva ultramétrica inducida sobre él, sería:

	1	2,3	4,5
1	0,0	0,8	1,0
2,3		0,0	1,0
4,5			0,0

Iteración 3ª:

Paso 1: Ahora, los 2 objetos más próximos serán los grupos $\{1\}$ y $\{2,3\}$ con un nivel de disimilaridad de 0,8

Paso 2: La nueva clasificación en 2 grupos sería $\{1,2,3\} \{4,5\}$

Paso 3: Y la nueva ultramétrica inducida sobre él, sería:

	1,2,3	4,5
1,2,3	0,0	1,0
4,5		0,0

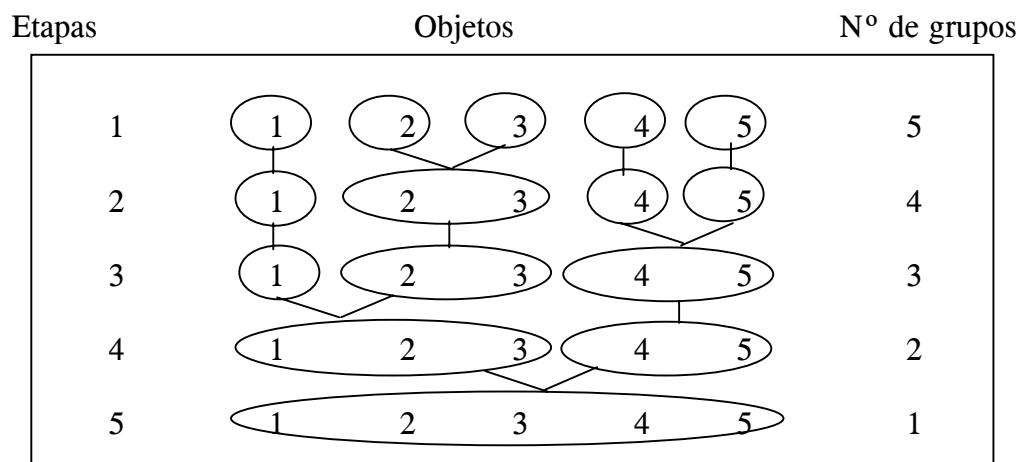
Iteración 4ª:

Paso 1: Ahora, los 2 objetos más próximos serán necesariamente los dos únicos grupos que tenemos de la iteración anterior $\{1,2,3\}$ y $\{4,5\}$ con un nivel de disimilaridad de 1,0

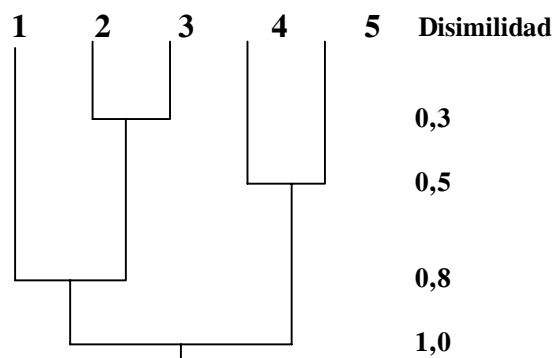
Paso 2: La nueva clasificación en 1 grupo sería obviamente en conjunto total $\{1,2,3,4,5\}$.

Paso 3: Ya que no hay más objetos que el grupo fusionado, no tiene sentido evaluar ninguna nueva ultramétrica inducida; con lo que finaliza el procedimiento de clasificación.

Resumidamente, el proceso de fusiones de objetos ha sido el siguiente:

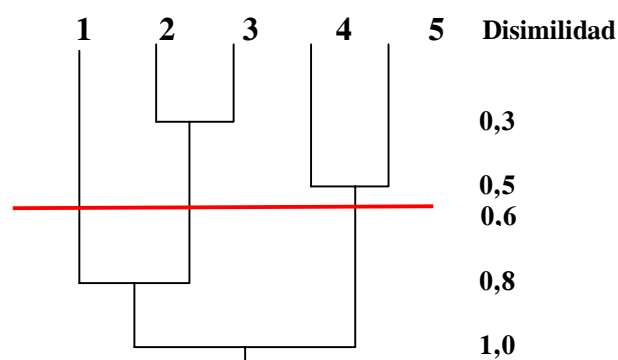


Lo que se suele representar, de forma más esquemática y añadiéndole la información del nivel de disimilaridad al que se realizan las distintas fusiones en el gráfico denominado *dendrograma*:



Así, por ejemplo, a una disimilaridad de 0.3 se unían los individuos 2 y 3; a una disimilaridad de 0.5 en la segunda iteración se unían los elementos 4 y 5; a 0.8 en la tercera, se unía el grupo 2-3 con el objeto 1; y, finalmente, a 1.0, en la cuarta, se unían los dos grupos resultantes que quedaban (1-2-3 y 4-5) para formar un único grupo constituido por todos los elementos: 1-2-3-4-5.

Si cortamos el *dendrograma* por un determinado nivel de disimilaridad, el número de líneas que se corten indicará el número de grupos que podemos obtener con una heterogeneidad interna máxima igual a dicho nivel. Así por ejemplo, si cortamos a un nivel de disimilaridad de 0.6,



observamos que, a ese nivel máximo de heterogeneidad interna tendremos tres grupos, pues cortamos 3 líneas verticales: un grupo formado sólo por el elemento 1, otro grupo formado por el 2 y 3 (cuya heterogeneidad es de 0,3), y un último grupo formado por los elementos 4 y 5 (cuya heterogeneidad interna es de 0,5).

Métodos Jerárquicos Aglomerativos

El algoritmo fundamental desarrollado en el apartado anterior produce clasificaciones jerárquicas únicas, claras y sin ambigüedades, siempre que la disimilaridad sea una ultramétrica. En este caso, para una ultramétrica, la definición de la disimilaridad de los objetos al nuevo grupo no tenía ningún problema ya que era común a todos los elementos del grupo y, consecuentemente, se elegía ese valor común como nueva medida de disimilaridad.

Sin embargo, las ultramétricas no se presentan frecuentemente en las situaciones reales; más bien, será prácticamente imposible que ocurra que todos los objetos a clasificar se encuentren dispuestos en triángulos isósceles. Y en consecuencia, el algoritmo fundamental se suele utilizar empleando cualquier medida de disimilaridad en lugar de la ultramétrica teóricamente exigida. Esto proporcionará clasificaciones jerárquicas diversas de acuerdo con las diferentes aproximaciones adoptadas para definir la disimilaridad de los objetos al nuevo grupo en función de las disimilaridades de aquéllos a los elementos de éste; lo que no era problema para la ultramétrica, por su unicidad, pero que ahora plantea diversas alternativas con distintas propiedades.

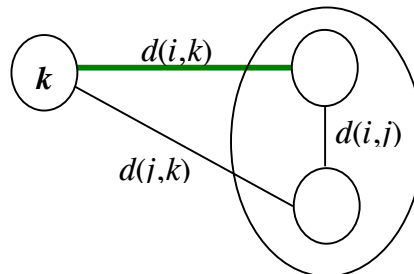
Podemos distinguir 3 familias de algoritmos: Métodos de enlace, Métodos basados en las varianzas o sumas de cuadrados de los grupos y Métodos basados en distancias entre centros de referencia de los grupos.

Método del Enlace Sencillo (o de la Distancia Mínima)

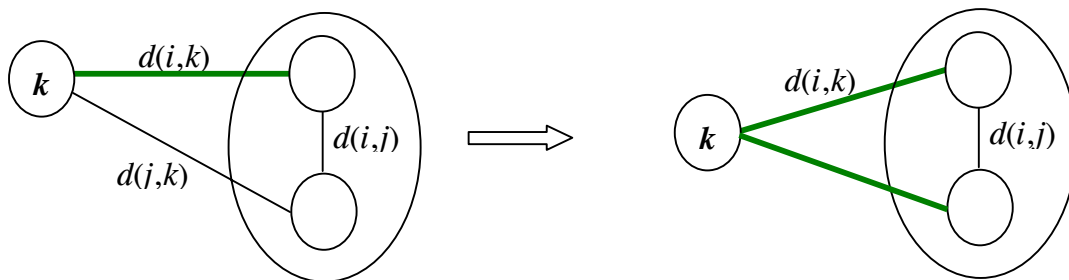
En este método, también llamado del enlace simple (simple linkage) o del vecino más próximo (nearest neighbour), calcula la disimilaridad de un objeto al nuevo grupo como la menor de las disimilaridades de aquél a los elementos del grupo; lo que se calcula como:

$$d(h_k, h_i \cup h_j) = \min\{d(h_k, h_i), d(h_k, h_j)\}$$

Si acabamos de unir los objetos i y j (por ser los que presentaban menor disimilaridad), podemos representar gráficamente la definición de la distancia de cualquier otro objeto k a este nuevo grupo que este método adopta, como sigue:

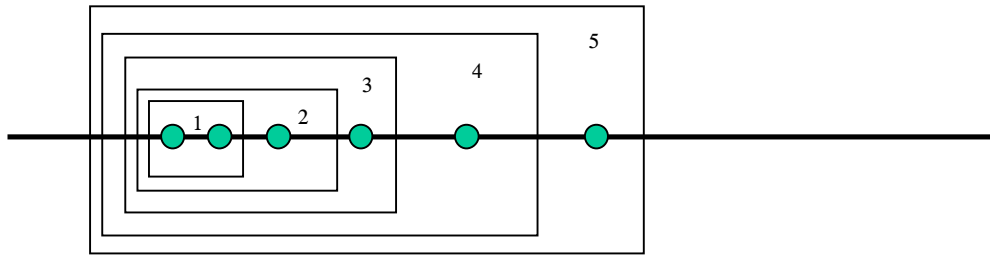


Notemos que, al adoptar esta definición como aproximación de la ultramétrica (en la que las disimilaridades a los elementos del nuevo grupo son iguales), estamos transformando un triángulo que no es isósceles en otro que sí lo es siendo sus lados mayores la distancia mínima calculada anteriormente.



Observemos que en cada iteración, la nube de puntos en el espacio multidimensional se está contrayendo. Así que, ya desde la segunda iteración, estamos clasificando unos objetos con una disposición espacial de acuerdo a sus disimilaridades que no es exactamente la inicial, por lo que las clasificaciones que hagamos se estarán viendo alteradas de alguna manera por esa contracción del espacio. Por ello, se dice que el método de la distancia mínima es de tipo *contractivo*.

Además, este método de la distancia mínima presenta el efecto llamado *encadenamiento*, por el que se favorece la formación de grupos “alargados”, como se representa en el gráfico siguiente:



en el que podemos observar que, en cada iteración, se iría agregando al grupo el objeto situado contiguamente al último objeto agregado, sin tener en cuenta que los primeros objetos quedan cada vez más lejanos.

Con este método, y debido a este efecto, se pueden producir grupos de tipo más bien alargados, que redondeados como los ejemplos del siguiente dibujo:



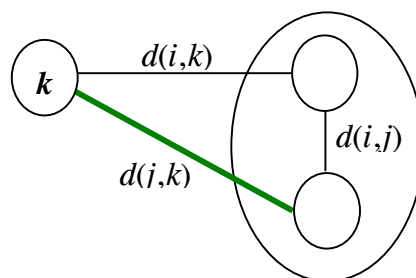
Método del Enlace Completo (o de la Distancia Máxima)

En este método, también llamado del enlace completo (complete linkage), calcula la disimilaridad de un objeto al nuevo grupo como la mayor de las disimilaridades de aquél a los elementos del grupo; lo que se calcula como:

$$d(h_k, h_i \cup h_j) = \text{Max}\{d(h_k, h_i), d(h_k, h_j)\}$$

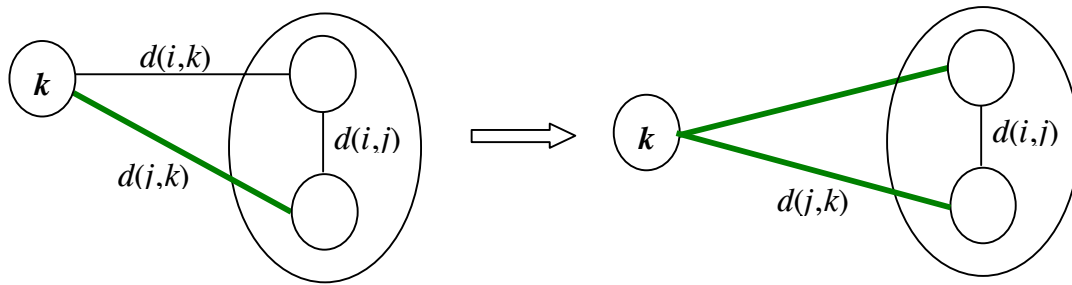
Si bien el método de la distancia mínima presentaba los efectos de contracción de la nube de puntos y encadenamiento de los objetos produciendo grupos alargados, este método actúa en sentido contrario, siendo más conservador en la exigencia de homogeneidad interna al agregar grupos, definiendo la distancia del nuevo grupo a un objeto externo es la mayor de las distancias a sus elementos.

Así, si acabamos de unir los objetos i y j (por ser los que presentaban menor disimilaridad), podemos representar gráficamente la definición de la distancia de cualquier otro objeto k a este nuevo grupo que este método adopta, como sigue:



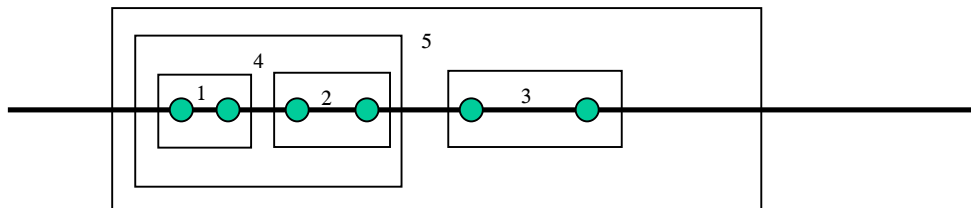
Y al adoptar esta definición como aproximación de la ultramétrica (en la que las disimilaridades a los elementos del nuevo grupo son iguales), estamos transformando un

triángulo que no es isósceles en otro que sí lo es siendo sus lados mayores la disimilaridad máxima calculada anteriormente.



Por tanto, a medida que se van sucediendo las iteraciones, estamos separando más los puntos de la nube y produciendo una cierta expansión de la misma respecto de su estructura en el espacio inicial; por lo que se dice que este método de la Distancia Máxima es de tipo *expansivo*.

Sin embargo, con relación a los grupos, su efecto será el opuesto al de la distancia mínima. Si aplicamos el método a los mismos datos allí presentados a modo de ejemplo, obtendríamos la siguiente secuencia de uniones,



donde se observa que el método produce en “engorde” paulatino de los grupos evitando que ningún elemento se aleje demasiado de ningún otro del mismo grupo; produciéndose así que los grupos están más redondeados, como los de la siguiente figura:



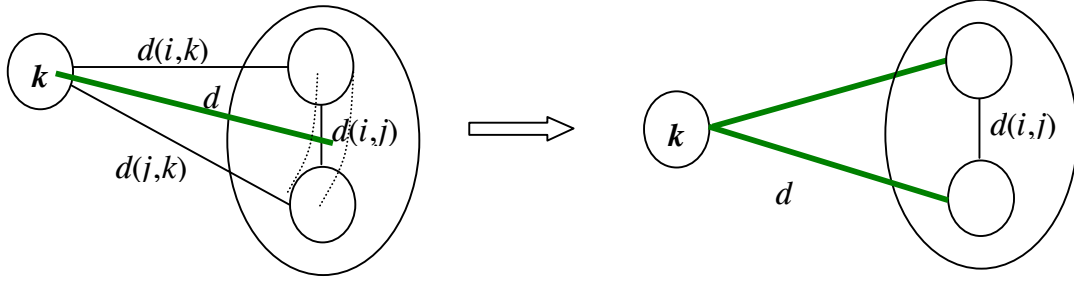
Como los dos métodos anteriores transforman la nube de puntos inicial bien contrayéndola (el de la distancia mínima) o bien dilatándolo (el de la distancia máxima), interesan otros métodos que conserven la estructura y dimensión de la nube de forma más realista; por lo que se desarrollan los métodos denominados *conservativos*. Vamos a ver a continuación algunos de estos métodos conservativos, que se diferencian entre sí sólo en la forma de definir la distancia del nuevo grupo agregado a cualquier objeto exterior al mismo, que en todos ellos se hará a partir de algún promedio de las disimilaridades iniciales.

Método del Enlace de la Distancia Media

Toma como nueva distancia entre el grupo k y el nuevo grupo $i \cup j$, el valor medio de las dos distancias

$$d(h_k, h_i \cup h_j) = \frac{d(h_k, h_i) + d(h_k, h_j)}{2}$$

Este valor medio gráficamente correspondería con una distancia que vendría aproximadamente por la línea de puntos en la siguiente representación:



Observemos que el método de la Distancia Media no tiene en cuenta los tamaños de los grupos que acaban de unirse. Es decir, las distancias a cada uno de los grupos a unir intervienen con la misma importancia, independiente de su cardinal; por lo que se está implícitamente sobreponderando a los grupos pequeños e infraponderando a los grandes.

Método UPGMA (Unweighted Pair Group Method using Arithmetic Average)

También conocido como Método del Enlace de la Distancia Media entre Grupos (Average Linkage Between Groups), este método UPGMA generaliza el método de la Distancia Media, pero ponderando las correspondientes distancias en función del tamaño de cada grupo para que todas las distancias entre cada dos elementos tengan la misma importancia en la evaluación de las distancias entre grupos. Como resultado, emplea la siguiente expresión para la nueva distancia entre el grupo k y el nuevo grupo $i \cup j$:

$$d(h_k, h_i \cup h_j) = \frac{1}{n_k(n_i + n_j)} \sum_{r \in h_k} \sum_{s \in h_i \cup h_j} d_{rs} = \frac{n_i}{n_i + n_j} d(h_k, h_i) + \frac{n_j}{n_i + n_j} d(h_k, h_j)$$

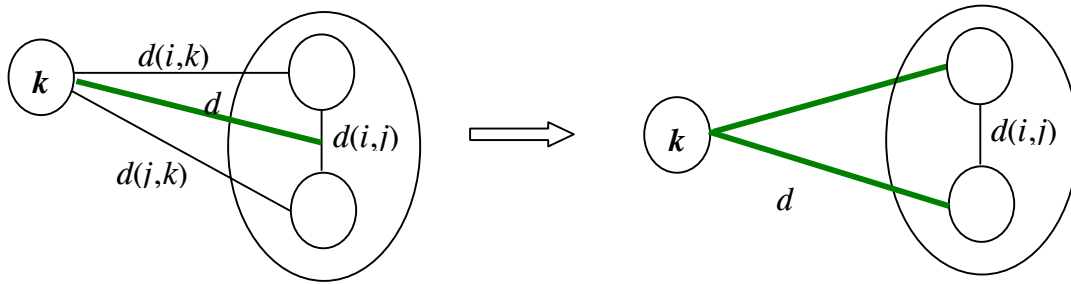
Así, si un grupo es más pequeño, su disimilaridad de referencia intervienen con menor peso que la del grupo mayor.

Método de la Mediana

Toma, como nueva distancia entre el grupo k y el nuevo grupo $i \cup j$, la longitud de la mediana del triángulo formado con los tres objetos i, j, k y trazada desde el objeto k . Esta longitud corresponde a la siguiente expresión:

$$d(h_k, h_i \cup h_j) = \frac{d(h_k, h_i)}{2} + \frac{d(h_k, h_j)}{2} - \frac{d(h_i, h_j)}{4}$$

lo que gráficamente puede interpretarse como sigue:



Observemos que, como el método de la Distancia Media, este método de la Mediana no tiene en cuenta los tamaños de los grupos que acaban de unirse. Es decir, las distancias a cada uno de los grupos a unir intervienen con la misma importancia, independiente de su cardinal; por lo que se está implícitamente sobreponderando a los grupos pequeños e infraponderando a los grandes. Por ello este método se generaliza, ponderando las correspondientes distancias de referencia entre grupos en función del tamaño de cada grupo resultando el método que se presenta a continuación.

Método del Centroide

Es una generalización del método de la mediana en el sentido anteriormente indicado; lo que conduce a la siguiente expresión para la nueva distancia entre el grupo k y el nuevo grupo $i \cup j$:

$$d(h_k, h_i \cup h_j) = \frac{n_i}{n_i + n_j} d(h_k, h_i) + \frac{n_j}{n_i + n_j} d(h_k, h_j) - \frac{n_i n_j}{(n_i + n_j)^2} d(h_i, h_j)$$

Así, si un grupo es más pequeño, su disimilaridad de referencia se tiene menos en cuenta, preponderando más la del grupo mayor.

Este método surge de la evaluación de la distancia entre grupos como el cuadrado de la distancia euclídea entre los centroides de cada grupo.

Fijémonos en que los cuatro métodos anteriores definen la nueva distancia como algún tipo de distancia promedio que cumple, por tanto, estar comprendida entre las dos distancias de referencia, produciéndose siempre un efecto conservativo sobre el espacio (nube de puntos) inicial, que, aunque se modifica un poco, prácticamente permanece con las mismas dimensiones que teníamos inicialmente.

Sin embargo, estos dos últimos métodos pueden presentar un problema: la ocurrencia de las llamadas inversiones. Podría ocurrir que en una determinada etapa se llegara a la decisión de unir dos objetos que podrían mostrar un nivel de disimilaridad menor que el presentado por otros objetos unidos en etapas anteriores; lo que es un contrasentido. Afortunadamente, cuando ocurre este efecto indeseable, suele hacerlo en las primeras etapas de fusión de grupos, no afectando generalmente a las clasificaciones finalmente aceptadas que suelen estar formadas por pocos grupos y consecuencia de alguna de las últimas etapas del algoritmo de clasificación.

Como ejemplo de que esto puede ocurrir, si $d(A,B)=0,383$, $d(A,C)=0,438$ y $d(B,C)=0,425$

obsérvese que el método de la mediana conduce a la distancia $d(A \cup B, C) = 0,33575 < d(A, B)$. Y con los mismos datos, y $n_A=3$, $n_B=4$, $n_C=5$, el método del Centroide conduce a la distancia $d(A \cup B, C) = 0,336 < d(A, B)$.

Método Flexible de Lance y Williams

Desde un punto de vista matemático, todos los métodos anteriores pueden considerarse derivados de uno sólo: del llamado *Método Flexible de Lance y Williams* que sintetizamos a continuación

$$d(h_k, h_i \cup h_j) = \alpha_i d(h_k, h_i) + \alpha_j d(h_k, h_j) + \beta d(h_i, h_j) + \gamma |d(h_k, h_i) - d(h_k, h_j)|$$

Como vemos, expresa la distancia del nuevo grupo agregado a cualquier objeto exterior como una expresión lineal de las distancias simples - $d(h_i, h_j)$, $d(h_k, h_i)$ y $d(h_k, h_j)$ - entre cada dos objetos de los 3 involucrados - i, j y k - en cada paso.

Según tomemos los valores de los parámetros α_i , α_j , β y γ obtendremos todos los métodos anteriores, tal y como se recoge en el siguiente cuadro:

Método	α_i	α_j	β	γ
Mínimo	1/2	1/2	0	-1/2
Máximo	1/2	1/2	0	1/2
Media	1/2	1/2	0	0
Mediana	1/2	1/2	-1/4	0
Centroide	$n_i / (n_i + n_j)$	$n_j / (n_i + n_j)$	$-n_i n_j / (n_i + n_j)^2$	0
UPGMA	$n_i / (n_i + n_j)$	$n_j / (n_i + n_j)$	0	0

Método del Enlace de la Distancia Media Intra Grupos (Average Linkage Within Group)

Este método es una variante del método UPGMA considerando como nueva distancia al grupo la distancia promedio de todos los elementos tomados dos a dos, independientemente de que sean del mismo grupo original o no. Como resultado, emplea la siguiente expresión para la nueva distancia entre el grupo k y el nuevo grupo $i \cup j$:

$$d(h_k, h_i \cup h_j) = \frac{1}{\left(\frac{n_k + (n_i + n_j)}{2} \right)} \sum_{r, s \in h_i \cup h_j \cup h_k} d_{rs}$$

Método de Ward (o de la mínima suma de cuadrados de distancias)

Este método calcula los centroides de cada grupo, \bar{x}_i , y evalúa la suma de los cuadrados de las distancias euclídeas de los elementos de cada grupo a su centroide como indicador de la homogeneidad/heterogeneidad de los elementos de cada grupo; y en consecuencia, une los grupos que producen un menor incremento de este indicador.

$$d(h_k, h_i \cup h_j) = \sum_{i \in h_i \cup h_j \cup h_k} d_2^2(i, \bar{x}_{h_i \cup h_j \cup h_k})$$

Todos los métodos vistos hasta ahora son métodos jerárquicos agregativos, lo que nos va a permitir describir el proceso de clasificación a través de un dendrograma (o alternativamente un *diagrama de témpanos o carámbanos*) a partir del cual se puede visualizar perfectamente el proceso de agregación de los objetos y de formación de la cadena jerárquica ente los grupos en las sucesivas iteraciones. Pasaremos a continuación a realizar un apunte sobre cómo podemos medir la transformación de la nube de puntos que producen estos métodos cuando no se emplea estrictamente una ultramétrica, para posteriormente revisar otros tipos de métodos de clasificación.

Evaluación del Proceso de Clasificación

Hemos visto que los métodos de clasificación jerárquica aglomerativos responden a un esquema teórico preciso que conducen a clasificaciones únicas, perfectamente claras, cuando partimos de ultramétricas. Pero generalmente en la práctica, no dispondremos de ultramétricas, sino de simples disimilaridades; por lo que la aplicación del algoritmo fundamental puede producir, en cada etapa o iteración, contracciones, expansiones o, en cualquier caso, transformaciones más o menos conservadoras de las dimensiones originales de la estructura de proximidades de los objetos a clasificar.

La pregunta pertinente es si después de las transformaciones sufridas tras el proceso iterativo, la estructura de proximidades que nos está sirviendo para clasificar a los objetos representa relativamente bien a la inicial.

Para comprobar hasta qué punto ocurre esto o no, podemos comparar la matriz de disimilaridades inicial, $((d_{ij}))$, con la matriz de las aproximaciones ultramétricas construida a lo largo del proceso iterativo, $((u_{ij}))$ que nos proporciona el dendrograma concreto que genera el método aplicado (disimilitud a la que se unen por primera vez).

$$u_{ij} = d(h), \text{ siendo } h \text{ el menor grupo que contiene a } i \text{ y } j$$

Así, comparando estas dos matrices, podemos de alguna manera ver si difieren sustancialmente, o no, las estructuras de proximidades inicial y final y, en consecuencia, si la clasificación es realista.

Para ello, se define el coeficiente de **correlación cofenética** como el coeficiente de correlación lineal de Pearson entre las disimilaridades iniciales y las ultramétricas finales. Así, si toma un valor próximo a 1, las alteraciones de la estructura de proximidades habrían

sido mínimas y, por tanto, la clasificación alcanzada sería bastante representativa para los datos iniciales. Cuanto más cercano a cero sea el valor de este coeficiente, más se estará alterando la estructura de proximidades original y, por tanto, menos realista será la clasificación obtenida.

Métodos Jerárquicos Divisivos

Como dijimos en la primera parte de este capítulo, los procedimientos jerárquicos divisivos, partiendo del grupo formado por todos los objetos, operan realizando particiones sucesivas de los grupos en función de alguna característica (métodos divisivos monotéticos) o de algunas características (métodos divisivos politéticos) que permita diferenciar a los elementos del grupo a dividir.

Por ejemplo, una clasificación como la siguiente es fruto de tal proceder.

$$\text{Seres} \left\{ \begin{array}{l} \text{Inanimados} \\ \text{Animados} \left\{ \begin{array}{l} \text{Animales} \left\{ \begin{array}{l} \text{Vertebrados} \\ \text{Invertebrados} \end{array} \right. \\ \text{Vegetales} \end{array} \right. \end{array} \right.$$

Esta clasificación jerárquica se derivan de observar, en cada paso, el comportamiento de una característica que distingue a los elementos del grupo a dividir. Así, en una primera etapa, observamos la característica de *animación* del ser, de forma que el grupo de seres se particiona en dos: los que poseen esa característica (grupo *Animados*) y los que no la tienen (grupo *Inanimados*). En la segunda etapa, observamos la característica *movilidad* de los seres animados, de forma que si la tienen, lo clasificamos como *Animal* y si no la tienen lo clasificamos como *Vegetal*. En la tercera etapa, tenemos en cuenta la *estructura* interna del ser, de forma que si posee un esqueleto óseo, diremos que es *Vertebrado* y si no, *Invertebrado*.

Las características que se han utilizado en este ejemplo son binarias; pero, en general, se pueden hacer clasificaciones similares con respuesta multinomial o múltiple. Al fin y al cabo, y como se vió en un capítulo anterior, las características multinomiales con k modalidades pueden representarse adecuadamente por la combinación de k característica binarias (o $k-1$ característica binarias independientes) referentes a la posesión o no de una modalidad determinada.

Por ejemplo, si tenemos una variable multinomial de 6 respuestas, la podemos modelizar como 5 variables binarias o dicotómicas como se indica en la tabla siguiente

X (Multinomial)	X ₁ (binaria)	X ₂ (binaria)	X ₃ (binaria)	X ₄ (binaria)	X ₅ (binaria)
0	0	0	0	0	0
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1

Método Williams y Lambert

El método divisivo de clasificación jerárquica de William y Lambert se emplea para clasificar objetos a partir de la observación de características binarias (dicotómicas).

El procedimiento es el general, ya introducido, para los métodos divisivos. Parte de considerar el conjunto de objetos como un todo; como un grupo único. Y fijada una característica binaria observada sobre los individuos del grupo, se particiona en subgrupos (en este caso en sólo 2 al ser la característica binaria) en función de que los elementos presenten (subgrupo 1º) o no presenten (subgrupo 2º) dicha característica; para volver a proceder de igual forma con otro grupo en la siguiente etapa y así sucesivamente.

Lo que caracteriza realmente a este método no es, pues, el procedimiento de división de grupos (que es el general), sino la elección de la característica que se empleará en cada etapa para que la homogeneidad dentro de los grupos y la heterogeneidad entre grupos sean las más adecuadas en algún sentido.

Así, para elegir cuál será la primera característica empleada para particionar el grupo, el método de William y Lambert compara las respuestas de los elementos del grupo ante cada variable binaria X_i con las obtenidas ante cada una de las otras variables X_j , recogiendo los resultados observados en una tabla de contingencia 2x2, T_{ij} característica de la comparación (X_i, X_j) , como la siguiente:

	X_j	
X_i	0	1
0	a	b
1	c	d

En esta tabla T_{ij} , a y d hacen referencia a los individuos del grupo que presentan comportamientos idénticos en las variables comparadas, mientras que c y b hacen referencia a los individuos del grupo que presentan comportamientos diferentes.

La medida de asociación por excelencia para tablas de contingencia es el coeficiente de contingencia χ^2 de Pearson, y que para la tabla T_{ij} notaremos por χ_{ij}^2 .

$$\chi_{i,j}^2 = \frac{N(ad - cb)^2}{(a+b)(a+c)(c+d)(b+d)}$$

A partir de esos coeficientes de contingencia, el método construye una matriz simétrica de similitudes entre variables (los χ_{ij}^2 crecen con la asociación, y cumplen que $\chi_{ij}^2 = \chi_{ji}^2$) como la representada a continuación:

	1	2	...	n
1	χ_{11}^2	χ_{12}^2	\dots	χ_{1n}^2
2		χ_{22}^2		\dots
\dots			\dots	\dots
n				χ_{nn}^2

El método toma como variable clasificadora en esta primera etapa, aquélla que hace máxima la suma de sus coeficientes de contingencia con las demás variables (es igual por filas que por columnas, dada la simetría de esta matriz):

$$\text{Max}_{i=1,\dots,p} \sum_{k=1}^p \chi_{ik}^2 = \text{Max}_{k=1,\dots,p} \sum_{i=1}^p \chi_{ik}^2$$

Es decir, el método elige la variable que presenta un mayor grado de asociación promedio con todas las demás, que presumiblemente contendrá mayor información común y redundante con el resto de variables y que permitirá discriminar más generalmente a los objetos. Y como consecuencia de esta primera etapa, el grupo se dividirá en dos en función de que los datos presenten o no esa característica.

Así, el proceso vuelve a reiniciarse de forma análoga en cada uno de los grupos creados, para obtener las sucesivas variables clasificadoras y subgrupos consecuentes con ellas.

Métodos No Jerárquicos

Uno de los resultados fundamentales del Análisis de la Varianza es que, cuando un conjunto de datos se particiona en un cierto número de grupos, k , la matriz de varianzas y covarianzas observada del conjunto original de datos, S , puede descomponerse como la suma de la llamada *matriz de varianzas y covarianzas intra-grupos*, W , más la llamada *matriz de varianzas y covarianzas inter-grupos*, B , según el siguiente resultado:

$$S = \frac{1}{n} X_c' X_c = B + W$$

siendo el elemento genérico de fila i y columna j de estas matrices, S , B y W respectivamente, los siguientes:

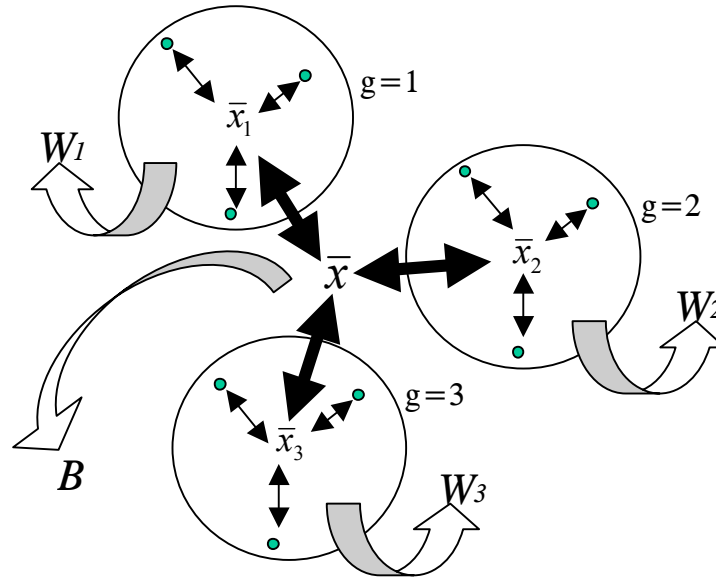
$$S_{ij} = \frac{1}{n} \sum_{g=1}^k \sum_{l=1}^{n_g} (x_{gl,i} - \bar{x}_i)(x_{gl,j} - \bar{x}_j)$$

$$B_{ij} = \frac{1}{n} \sum_{g=1}^k n_g (\bar{x}_{g,i} - \bar{x}_i)(\bar{x}_{g,j} - \bar{x}_j)$$

$$W_{ij} = \frac{1}{n} \sum_{g=1}^k \sum_{l=1}^{n_g} (x_{gl,i} - \bar{x}_{g,i})(x_{gl,j} - \bar{x}_{g,j})$$

donde n representa el número de datos; X_c , la matriz de los datos centrados; \bar{x}_i , la media de la variable i en el conjunto de datos; n_g , el número de elementos del grupo g -ésimo; $\bar{x}_{g,i}$, la media de la variable i en el grupo g -ésimo; y $x_{gl,i}$, el valor que presenta el individuo l -ésimo del grupo g -ésimo para la variable i .

Lo que gráficamente puede esquematizarse en el siguiente gráfico:



La dispersión dentro de cada grupo produce una matriz de varianzas y covarianzas W_i , indicativa de la homogeneidad interna en los grupos. Y sumando éstas para todos los grupos se obtendría la matriz de varianzas y covarianzas intra-grupos W , indicativa de la homogeneidad interna de los grupos en conjunto.

Por su parte, la matriz B o matriz de varianzas inter-grupos, se obtendría como varianza de los centroides de los grupos con respecto al centroide de todos los datos en conjunto, ponderando aquéllos por el número de elementos del grupo al que corresponde y representa. Por tanto, la varianza inter-grupos es indicativa de la separación o heterogeneidad entre los grupos.

Por tanto, a medida que una determinada distribución en grupos de los objetos “aumenta” en algún sentido la matriz W , también hará disminuir en algún sentido la matriz B , ya que el conjunto de objetos, sus datos y la correspondiente matriz de varianzas y covarianzas S permanecen inalterados; y viceversa.

Los métodos de clasificación no jerárquicos suelen utilizar, de una u otra forma, esta propiedad. Buscarán clasificaciones que separen lo más posible los grupos (“aumenten” en algún sentido la matriz B), con lo que consecuentemente consiguen una mayor homogeneidad interna de éstos (“disminuyen” consecuentemente la matriz W).

Para ello, prefijan el número k de grupos en que se pretenden clasificar a los elementos, reparten inicialmente los elementos a clasificar en k grupos provisionales a partir de los que empezar a trabajar y, mediante procedimientos de optimización, reasignan a los elementos a uno u otro grupo hasta obtener una configuración óptima en el sentido anteriormente

indicado.

La asignación provisional inicial en los k grupos, puede hacerse, según el método de muy diferentes formas. La más usual es distribuir en el espacio k puntos, que actuarán como centroides de los grupos iniciales, (uniformemente, en forma de simplex, subjetivamente, etc) y asignar al grupo representado por cada uno de estos puntos a los elementos que se encuentren más cercanos a él que a ningún otro centroide.

Por otro lado, la búsqueda de la distribución óptima de puntos en grupo, de forma que los grupos sean internamente lo más homogéneos posible y que las separaciones entre grupos sean lo mayores posibles, se puede realizar mediante alguna función objetivo que exprese el que “ B sea máxima” y “ W se mínima”, en algún sentido. Dado que B y W no son escalares sino matrices, el objetivo que perseguimos se consigue, entre otras, con funciones objetivos como las siguientes:

Minimizar la Traza (W)
 Minimizar el Determinante (W)
 Minimizar el cociente $L = \text{Determinante}(W) / \text{Determinante}(S)$
 Maximizar la Traza ($W^T B$)

Estos procedimientos, aplicados a valores distintos de k , producen generalmente sendas clasificaciones no relacionadas jerárquicamente, ya que nada garantiza que elementos que se asignen unidos en un mismo grupo en una clasificación de k grupos, permanecerían unidos igualmente cuando se plantease la clasificación con un menor número de grupos.

Método K-medias

Este método plantea la minimización de la suma de las distancias euclídeas de los elementos a los centros de los grupos (nótese que, cuando los centros son los centroides de los grupos, esto es equivalente a minimizar la suma de varianzas intra-grupos de todas las variables).

Y para ello combina dos algoritmos:

Un algoritmo de selección de centroides iniciales
 Un algoritmo de reasignación óptima de grupos

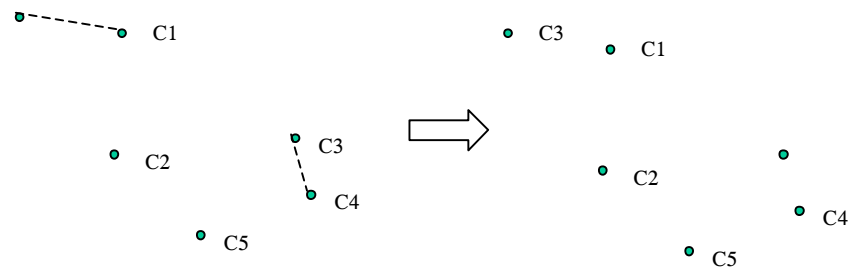
Algoritmo de selección de centroides iniciales

Procurando obtener una configuración de elementos (centros iniciales) lo más separados posible, realiza los siguientes pasos:

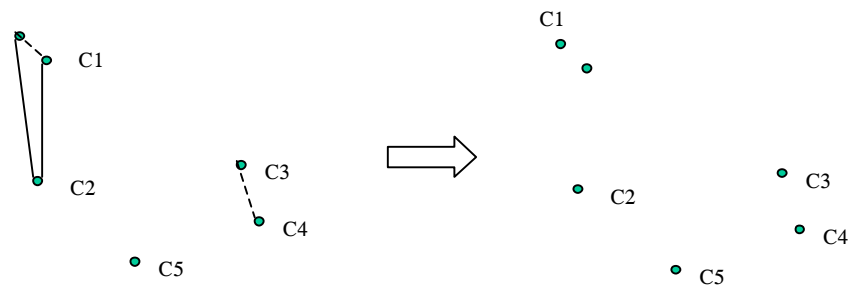
Paso 1. Se seleccionan los k primeros casos como centros provisionales

Paso 2. Se considera un nuevo punto.

a) Si éste dista de su centro más cercano más de lo que distan entre sí los dos centros más próximos, pasa a reemplazar como centro, al que de éstos se encontrase más cercano de él.



b) Si no fuera así, pero el nuevo punto distara más del segundo centro más cercano que la distancia más corta de los demás centros al más cercano del nuevo punto, entonces el punto reemplazaría a este centro más cercano.



Paso 3. Se reitera el proceso desde el paso 2 hasta que no haya más puntos que considerar

Como consecuencia, se obtendrá una configuración de centros constituida por los puntos más distantes entre sí de la nube de puntos a clasificar.

Algoritmo de reasignación óptima de grupos

Procurando obtener una configuración óptima de grupos lo más homogéneos posible internamente (lo que asegura que estarán lo más separados posible), realiza los siguientes pasos:

Paso 0. Se parte de k centros (especificados o calculados)

Iteración h ($h=1,2,\dots$):

Paso 1. Se calculan las distancias euclídeas de cada punto a cada centro

Paso 2. Se asigna cada punto al grupo representado por el centro que se encuentra a una menor distancia de él.

Paso 3. Se recalculan los nuevos centros como los centroides de los grupos

Paso 4. Terminar si no se aprecian cambios en los centros de los grupos con respecto de la iteración anterior o si se alcance alguna regla de paro del algoritmo (máximo número de iteraciones alcanzado, desplazamiento inapreciable de centroides, etc).

En caso contrario, se reitera el proceso desde el paso 1