Para el caso de datos continuos, se tienen las distancias provenientes de las normas en L^r; $r \geq 1$

$$d_{ij} = \|x_{G_i} - x_{G_j}\|_r = \left\{ \sum_{k=1}^{p} |x_{ik} - x_{jk}|^r \right\}^{1/r}, \quad \ldots\ldots (i)$$

estas distancias miden disimilaridad entre individuos.

Al calcular distancias entre individuos con (i), se esta asumiendo en forma implicita que las variables (las componentes de x_G) fueron medidas en la misma escala. Si este no fuera el caso, entonces se puede aplicar una "estandarización". Se considera una matriz A que sea positivo definida y de dimensiones $p \times p$ para definir

$$d_{ij}^2 = \|x_{G_i} - x_{G_j}\|_A = (x_{G_i} - x_{G_j})' A (x_{G_i} - x_{G_j}),$$

por ejemplo la norma en L^2 se obtiene si $A = I_p$, pero en el caso de requerir una estandarización, se puede usar

$$A = \text{diag}\left(\frac{1}{\widehat{VAR}(X_1)}, \frac{1}{\widehat{VAR}(X_2)}, \ldots, \frac{1}{\widehat{VAR}(X_p)} \right)$$

de esta forma $d_{ij}^2 = \sum_{k=1}^{p} \left\{ \frac{(x_{ik} - x_{jk})^2}{\widehat{VAR}(X_k)} \right\}$. Este ejemplo es importante, porque se evita que las distancias dependan del tipo de unidades de medición.

ALGORITMOS PARA CONGLOMERADOS

Esencialmente hay dos tipos de algoritmos para formar grupos: algoritmos jerárquicos y algoritmos de particiones. A su vez, los algoritmos jerárquicos se pueden dividir en procedimientos aglomerativos y procedimientos separativos. El algoritmo jerárquico aglomerativo comienza por la partición más fina posible (aquella en la que cada grupo ó subconjunto tiene una sola observación o individuo) y con esta propone nuevos grupos con estructura más compleja. El algoritmo jerárquico separativo, comienza con la partición menos fina posible (aquella en la que sólo hay un grupo, el cual contiene a todas las observaciones o individuos) y procede a

separar este conglomerado en grupos más pequeños

Estos algoritmos trabajan intercambiando elementos entre grupos hasta que logran optimizar una función score. La diferencia principal entre los algoritmos jerárquicos y los algoritmos de particiones, consiste en que para los algoritmos jerárquicos una vez que se encuentra la estructura de conglomerados "óptima", esta ya no se puede modificar. Para los algoritmos de particiones si es posible modificar la estructura de conglo-merados

ALGORITMOS JERARQUICOS (AGLOMERATIVOS)

Algoritmo:

1 Construir la partición más fina

2 Calcular la matriz de distancias D

3 Repetir:

(4) Encontrar aquellos dos grupos con distancia más pequeña entre ellos

(5) Formar un grupo con los dos grupos seleccionados.

(6) Calcular las distancias entre los nuevos grupos y construir una matriz de distancias (reducida) D

7 Detener el proceso cuando el único grupo que resulte es el de todas las observaciones $\mathcal{X}$.

Si dos individuos o grupos P y Q forman un nuevo grupo, para calcular la distancia entre este nuevo grupo y el grupo R se utiliza la siguiente distancia ponderada (Lance y Williams)

$$d(R,\{P,Q\}) = \delta_1 d(R,P) + \delta_2 d(R,Q) + \delta_3 d(P,Q) + \delta_4 |d(R,P) - d(R,Q)|$$

Los pesos $\delta_1, \delta_2, \delta_3$ y δ_4 definen diferentes distancias ponderadas (diferentes algoritmos aglomerativos), para mencionar algunos ejemplos sean $n_P \equiv \sum_{i=1}^{n} \mathbb{1}_{[x_i \in P]}$, $n_Q \equiv \sum_{i=1}^{n} \mathbb{1}_{[x_i \in Q]}$ y $n_R = \sum_{i=1}^{n} \mathbb{1}_{[x_i \in R]}$ el número de elementos en P, Q y R,

Table 4.1 Standard agglomerative hierarchical clustering methods.

Method	Alternative name[a]	Usually used with:	Distance between clusters defined as:	Remarks
Single linkage Sneath (1957)	Nearest neighbour	Similarity or distance	Minimum distance between pair of objects, one in one cluster, one in the other	Tends to produce unbalanced and straggly clusters ('chaining'), especially in large data sets. Does not take account of cluster structure.
Complete linkage Sorensen (1948)	Furthest neighbour	Similarity or distance	Maximum distance between pair of objects, one in one cluster, one in the other	Tends to find compact clusters with equal diameters (maximum distance between objects). Does not take account of cluster structure.
(Group) Average linkage Sokal and Michener (1958)	UPGMA	Similarity or distance	Average distance between pair of objects, one in one cluster, one in the other	Tends to join clusters with small variances. Intermediate between single and complete linkage. Takes account of cluster structure. Relatively robust.
Centroid linkage Sokal and Michener (1958)	UPGMC	Distance (requires raw data)	Squared Euclidean distance between mean vectors (centroids)	Assumes points can be represented in Euclidean space (for geometrical interpretation). The more numerous of the two groups clustered dominates the merged cluster. Subject to reversals.
Weighted average linkage McQuitty (1966)	WPGMA	Similarity or distance	Average distance between pair of objects, one in one cluster, one in the other	As for UPGMA, but points in small clusters weighted more highly than points in large clusters (useful if cluster sizes are likely to be uneven).
Median linkage Gower (1967)	WPGMC	Distance (requires raw data)	Squared Euclidean distance between weighted centroids	Assumes points can be represented in Euclidean space for geometrical interpretation. New group is intermediate in position between merged groups. Subject to reversals.
Ward's method Ward (1963)	Minimum sum of squares	Distance (requires raw data)	Increase in sum of squares within clusters, after fusion, summed over all variables	Assumes points can be represented in Euclidean space for geometrical interpretation. Tends to find same-size, spherical clusters. Sensitive to outliers.

[a]U = unweighted; W = weighted; PG = pair group; A = average; C = centroid.

entonces tenemos

TABLA 2

Nombre de la distancia	δ_1	δ_2	δ_3	δ_4
Single Linkage	$1/2$	$1/2$	0	$-\frac{1}{2}$
Complete Linkage	$1/2$	$1/2$	0	$1/2$
Average Linkage (unweighted)	$1/2$	$1/2$	0	0
Average Linkage (weighted)	$\frac{n_P}{n_P+n_Q}$	$\frac{n_Q}{n_P+n_Q}$	0	0
Centroid	$\frac{n_P}{n_P+n_Q}$	$\frac{n_Q}{n_P+n_Q}$	$-\frac{n_P n_Q}{(n_P+n_Q)^2}$	0
Median	$1/2$	$1/2$	$-1/4$	0
Ward	$\frac{n_R+n_P}{n_R+n_P+n_Q}$	$\frac{n_R+n_Q}{n_R+n_P+n_Q}$	$-\frac{n_R}{n_R+n_P+n_Q}$	0

Para el caso de Single Linkage y Complete Linkage, se suelen usar los siguientes algoritmos aglomerativos modificados

ALGORITMO (JERARQUICO) AGLOMERATIVO MODIFICADO

1 construir la partición más fina.

2 Calcular la matriz de distancias D

3 Repetir:

(4) Encontrar el mínimo (Single Linkage) / máximo (Complete Linkage) valor de d (entre dos individuos n y m) en D