5) Si m y n no están en el mismo grupo, entonces se forma un nuevo grupo con estos y se elimina el valor de d encontrado en 4

6) Construir la matriz de distancias con los nuevos grupos

7 Detener el proceso cuando:

(a) Todos los grupos fueron aglomerados en $\mathcal{X}$

(b) El valor de d satisface un criterio preestablecido

Representación Gráfica de la sucesión de agrupamientos

DENDOGRAMA

Una gráfica muy usada para representar la sucesión de agrupamientos es el dendograma. Esta gráfica despliega las observaciones, la sucesión de grupos producida por el algoritmo de agrupamiento y las distancias entre los grupos. En el eje horizontal se suelen representar a los individuos con índices y en el eje vertical se representan las distancias entre los grupos (en algunos paquetes en el eje horizontal se representan distancias y en el eje vertical a los individuos). El dendograma es similar a un árbol que "ramifica" conforme las distancias se hacen pequeñas, al trazar una linea horizontal y cortar el árbol a una determinada distancia, las ramas describen una estructura de agrupamientos.

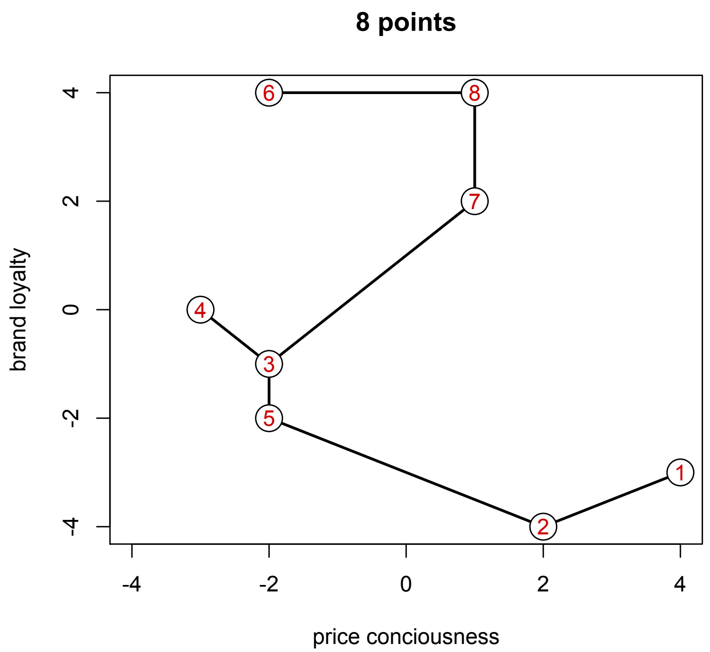
Ejemplo: Se tiene una muestra de 8 puntos

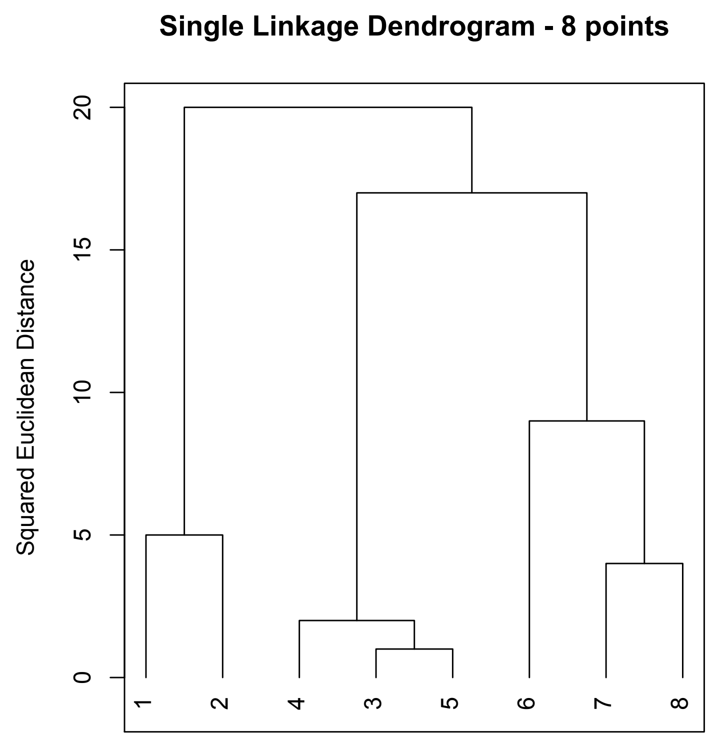
$x_1, \ldots, x_8 \qquad x_i' = (X_1, X_2)$

X_1 = Conciencia del Precio $\quad X_2$ = Lealtad a la marca

VÉASE FIGURA A

Supóngase que se lleva a cabo un algoritmo aglomerativo usando single Linkage. La matriz de distancias (normas L^2) está dada por


8 points
brand loyalty
price conciousness
4
2
0
-2
-4
-4
-2
0
2
4
6
8
7
4
3
5
1
2


Single Linkage Dendrogram - 8 points
Squared Euclidean Distance
20
15
10
5
0
1
2
4
3
5
6
7
8

$$D = \begin{pmatrix} 0 & 10 & 53 & 73 & 50 & 98 & 41 & 65 \\ & 0 & 25 & 41 & 20 & 80 & 37 & 65 \\ & & 0 & 2 & 1 & 25 & 18 & 34 \\ & & & 0 & 5 & 17 & 20 & 32 \\ & & & & 0 & 36 & 25 & 45 \\ & & & & & 0 & 13 & 9 \\ & & & & & & 0 & 4 \\ & & & & & & & 0 \end{pmatrix}$$

Al aplicar el algoritmo aglomerativo se obtiene el dendograma en la figura A. Si por ejemplo el arbol se corta a nivel 10, tenemos una estructura de 3 grupos: $\{1,2\}$, $\{3,4,5\}$ y $\{6,7,8\}$

El algoritmo single Linkage, define la distancia entre dos grupos como el mínimo de las distancias individuales (las distancias entre individuos de los grupos), de la Tabla 2 (página 10) tenemos

$$d(R,\{P,Q\}) = \tfrac{1}{2}d(R,P) + \tfrac{1}{2}d(R,Q) - \tfrac{1}{2}\left|d(R,P) - d(R,Q)\right|$$

<u>caso 1</u>: $d(R,P) \geq d(R,Q)$

$$d(R,\{P,Q\}) = \tfrac{1}{2}d(R,P) + \tfrac{1}{2}d(R,Q) - \tfrac{1}{2}\left(d(R,P)\right) + \tfrac{1}{2}\left(d(R,Q)\right)$$

$$= d(R,Q) = \min\left(d(R,P), d(R,Q)\right)$$

Caso 2 $\quad d(R,P) \leq d(R,Q)$

$$d(R,\{P,Q\}) = \frac{1}{2}d(R,P) + \frac{1}{2}d(R,Q) - \frac{1}{2}\left(d(R,Q) - d(R,P)\right)$$
$$= d(R,P) = \min\left(d(R,P), d(R,Q)\right)$$

$$\therefore \quad d(R,\{P,Q\}) = \min\left(d(R,P), d(R,Q)\right) \quad \ldots\ldots \text{(a)}$$

Por esta razón a este algoritmo también se le llama "algoritmo del vecino más cercano".

Como consecuencia de esta característica, el algoritmo single Linkage tiende a construir grupos numerosos. Puede suceder que los grupos no resulten "muy distantes" en el sentido de que existen puntos o individuos que esten "cercanos" pese a que estos pertenecen a diferentes grupos, lo anterior puede ocasionar que el algoritmo reuna a los correspondientes grupos en uno solo.

El algoritmo complete Linkage trata de corregir el problema anterior al considerar el máximo de las distancias individuales

Se puede argumentar, igual que como se hizo para la ecuación (a) en el single Linkage, que para el caso del algoritmo complete Linkage

$$d(R,\{P,Q\}) = \max\left\{d(R,P),\, d(R,Q)\right\},$$

razón por la cual también se le conoce como el "algoritmo del vecino más lejano"