几个常用来衡量聚类间的相近程度公式，说明如下：

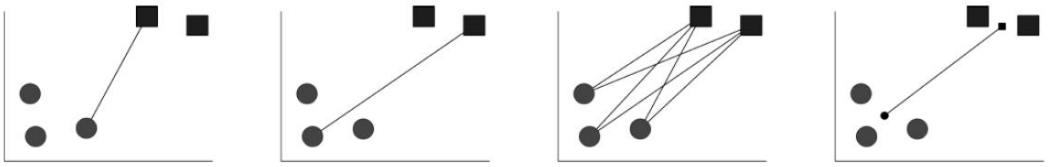
$$\text{最小距离(minimum distance)：} D_{\min}(C_i, C_j) = \min_{a \in C_i, b \in C_j} D_{(a,b)} \tag{6.10}$$

$$\text{最大距离(maximum distance)：} D_{\max}(C_i, C_j) = \max_{a \in C_i, b \in C_j} D_{(a,b)} \tag{6.11}$$

$$\text{平均距离(average distance)：} D_{\text{average}}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{a \in C_i} \sum_{b \in C_j} D_{(a,b)} \tag{6.12}$$

$$\text{中心值距离(centroid distance)：} D_{\text{centroid}}(C_i, C_j) = D_{(m_i, m_j)} \tag{6.13}$$

其中，m_i 与 m_j 分别表示聚类 C_i 与 C_j 的中心值，n_i 与 n_j 分别表示聚类 C_i 与 C_j 的数据点个数，$D_{(a,b)}$ 表示两样本点间的距离，可以使用的距离衡量方式有欧式距离或曼哈顿距离等，更进一步的说明可见图 6.3。



(a) 最小距离　(b) 最大距离　(c) 平均距离　(d) 中心值距离

图 6.3　距离示意图

[範例 6.1]為 7 筆觀察值的 V1 與 V2 資料(如表 6.2),為方便計算,以歐式距離平方作為衡量相似度的依據,可計算出各資料點間的歐式距離平方如表 6.3 所列。假設現在有三個聚類,分別是聚類 A={1,3,6},聚類 B={2,4},聚類 C={5,7},聚類 A 與 B 間共有 6 個距離,分別為:$D_{1\&2}$=233、$D_{1\&4}$=261、$D_{3\&2}$=149、$D_{3\&4}$=169、$D_{6\&2}$=80、$D_{6\&4}$=104.

表 6.2[範例 6.1]觀察值

觀察值	V1	V2
Y1	14	15
Y2	22	28
Y3	15	18
Y4	20	30
Y5	30	35
Y6	18	20
Y7	32	30

歐氏距離平方

序號	1	2	3	4	5	6	7
1	0	233	10	261	656	41	549
2	233	0	149	8	113	80	104
3	10	149	0	169	514	13	433
4	261	8	169	0	125	104	144
5	656	113	514	125	0	369	29
6	41	80	13	104	369	0	296
7	549	104	433	144	29	296	0

最小距離： $D_{min(C_A,C_B)}=D_{6\&2}=80$

最大距離： $D_{max(C_A,C_B)}=D_{1\&4}=261$

平均距離： $D_{average(C_A,C_B)}=\frac{D_{1\&2}+D_{1\&4}+D_{3\&2}+D_{3\&4}+D_{6\&2}+D_{6\&4}}{6}=166$

中心值距離： *聚類A的中心*：$\left(\frac{14+15+18}{3}\quad\frac{15+18+20}{3}\right)=\left(\frac{47}{3}\quad\frac{53}{3}\right)$

聚類B的中心：$\left(\frac{22+20}{2}\quad\frac{28+30}{2}\right)=(21,29)$

則，聚類 A 與聚類 B 的歐氏距離為：

$$D_{centroid(C_A,C_B)}=\left(21-\frac{47}{3}\right)^2+\left(29-\frac{53}{3}\right)^2=156.89$$

常見的層次聚類分析方法包括:單一連結法(single linkage method),以兩聚類間資料點中的最小距離來表示兩聚類的距離及兩群資料的鄰近程度;完全連結法(complete linkage method),以兩聚類間資料點的最大距離來表示兩聚類的距離及兩群資料的鄰近程度;平均連結法(average linkage method),衡量聚類內所有點到另一個聚類內所有點的距離平均來表示兩聚類的鄰近程度,以避免聚類之間的距離衡量受雜訊影響;中心點連結法(centroid linkage method),以兩聚類的中心點距離作為衡量兩聚類的距離,以表示其鄰近程度。

以[範例 6.1]為例,利用單一連結法說明層次聚類分析的計算,起初所有資料皆屬於單一聚類,而資料點 2 與資料點 4 最接近,所以將兩點合併為一聚類,重新計算各聚類間資料點的最小距離如表 6.4 所示。而資料點 1 與資料點 3 最為接近,因此將兩點合併為新的聚類，迭代，直到將所有數據點均合併至同一聚類中為止。

單一連結法，合併 2 和 4 後的歐氏距離

序號	1	2&4	3	5	6	7
1	0	233	10	656	41	549
2&4	233	0	149	113	80	104
3	10	149	0	514	13	433
5	656	113	514	0	369	29
6	41	80	13	369	0	296
7	549	104	433	29	296	0

最後聚類 AB 與聚類 C 在距離為 104 時合併為一群

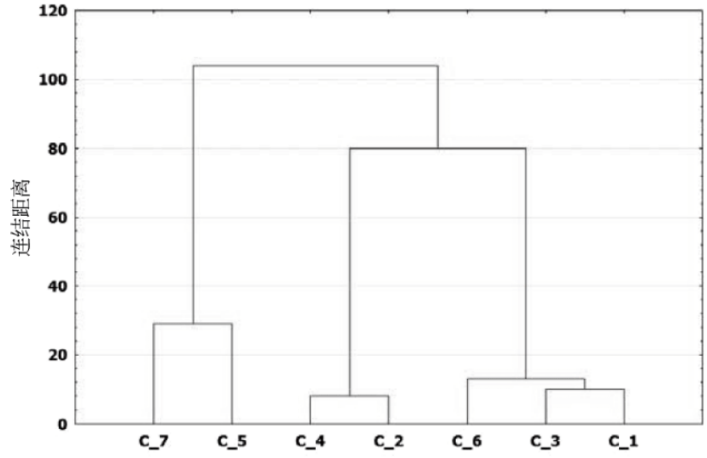


图 6.4　单一连结法树形图