

خوشه‌بندی سلسله‌مراتبی

Hierarchical Clustering

خوشه‌بندی $\mathcal{C} = \{C_1, \dots, C_k\}$ یعنی افراز دسته‌داده‌ی D به k زیر گروه $C_i \subseteq D$

$$C_i \cap_{i \neq j} C_j = \emptyset$$

$$\bigcup_{i=1}^k C_i = D$$

خوشه‌بندی آشیانه‌ای یا تودرتو: خوشه‌بندی $A = \{A_1, \dots, A_r\}$ در خوشه‌بندی $B = \{B_1, \dots, B_s\}$ آشیانه کرده است اگر $r > s$ و برای هر $A_i \in A$ یک خوشه‌ی $B_j \in B$ وجود داشته که $A_i \subseteq B_j$ باشد.

خوشه‌بندی سلسله‌مراتبی همواره منجر به توالی از n خوشه‌بندی آشیانه‌ای C_1, \dots, C_n می‌شود.

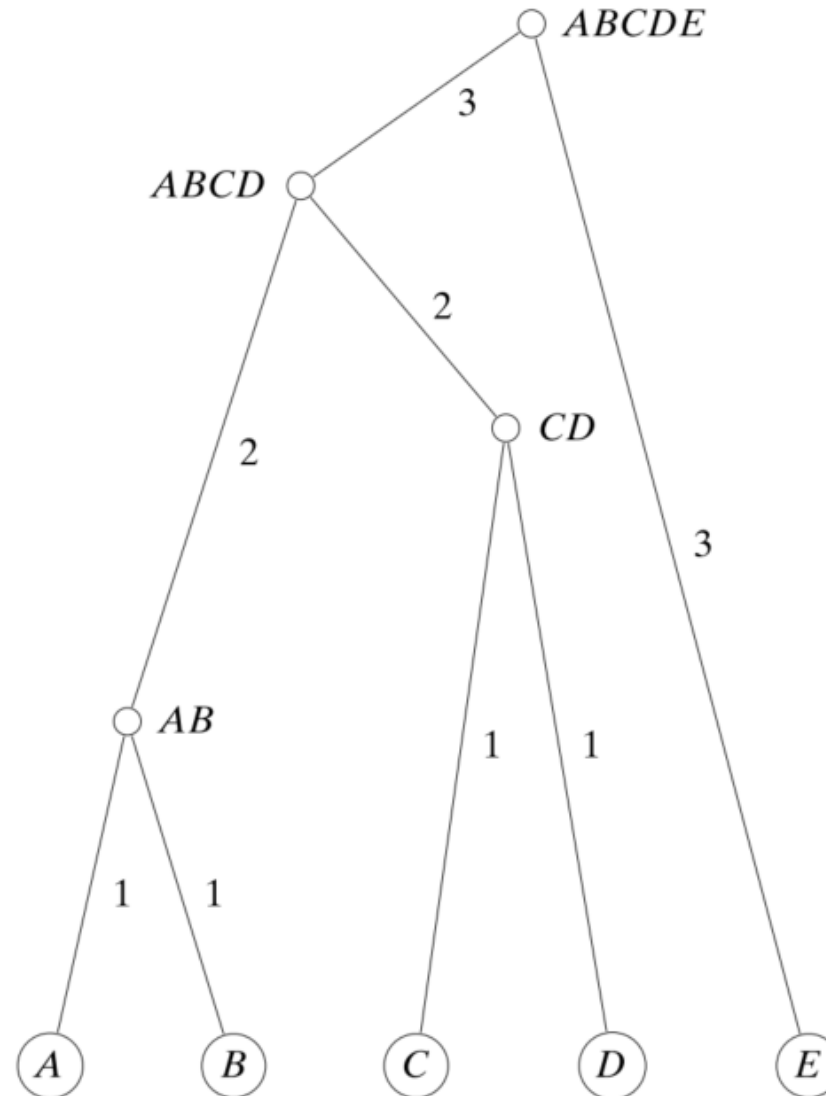
که از خوشه‌بندی با خوشه‌های تک عضوی $C_1 = \{\{x_1\}, \dots, \{x_n\}\}$ شروع و

به خوشه‌بندی با خوشه‌ی n عضوی $C_n = \{\{x_1, \dots, x_n\}\}$ خاتمه می‌یابد.

در حالت عمومی خوشه‌ی C_{t-1} در خوشه‌ی C_t آشیانه کرده است. یعنی اگر $C_i \in C_{t-1}$ و $C_j \in C_t$ باشد، آنگاه $C_i \subset C_j$

یک از روش‌های محاسبه‌ی فاصله بین خوشه‌ها روش تک پیوند (Single Link) است.

$$\delta(C_i, C_j) = \min\{\|\mathbf{x} - \mathbf{y}\| \mid \mathbf{x} \in C_i, \mathbf{y} \in C_j\}$$



δ	E
$ABCD$	3

δ	CD	E
AB	2	3
CD		3

δ	C	D	E
AB	3	2	3
C		1	3
D			5

δ	B	C	D	E
A	1	3	2	4
B		3	2	3
C			1	3
D				5

Figure 14.3. Single link agglomerative clustering.

Algorithm 14.1: Agglomerative Hierarchical Clustering Algorithm

AGGLOMERATIVECLUSTERING(\mathbf{D}, k):

- 1 $\mathcal{C} \leftarrow \{C_i = \{\mathbf{x}_i\} \mid \mathbf{x}_i \in \mathbf{D}\}$ // Each point in separate cluster
 - 2 $\Delta \leftarrow \{\|\mathbf{x}_i - \mathbf{x}_j\| : \mathbf{x}_i, \mathbf{x}_j \in \mathbf{D}\}$ // Compute distance matrix
 - 3 **repeat**
 - 4 Find the closest pair of clusters $C_i, C_j \in \mathcal{C}$
 - 5 $C_{ij} \leftarrow C_i \cup C_j$ // Merge the clusters
 - 6 $\mathcal{C} \leftarrow (\mathcal{C} \setminus \{C_i, C_j\}) \cup \{C_{ij}\}$ // Update the clustering
 - 7 Update distance matrix Δ to reflect new clustering
 - 8 **until** $|\mathcal{C}| = k$
-