

Übungsblatt 5

04.07.2014

Aufgabe 5.1:

- (a) Geben Sie eine Distanzmatrix an, welche metrisch, aber nicht additiv ist.
- (b) Geben Sie eine Distanzmatrix an, welche additiv, aber nicht ultrametrisch ist.

Aufgabe 5.2:

Gegeben seien folgende Distanzmatrizen:

a)	1	2	3	4	5
1	0				
2	8	0			
3	8	2	0		
4	12	12	12	0	
5	8	4	4	12	0

b)	1	2	3	4	5
1	0				
2	3	0			
3	7	8	0		
4	10	11	9	0	
5	7	8	6	5	0

Welche der Matrizen induziert eine Ultrametrik? (Mit Beweis!)

Konstruieren Sie ggf. den zugehörigen ultrametrischen Baum mit Hilfe des in der Vorlesung vorgestellten Algorithmus.

Aufgabe 5.3:

Der Algorithmus UPGMA wurde ebenfalls in der Vorlesung vorgestellt und ist auf der nächsten Seite noch einmal zusammengefaßt.

- a) Wenden Sie UPGMA auf eine der beiden Distanzmatrizen aus Aufgabe 5.2 an (d.h. konstruieren Sie schrittweise den entsprechenden Baum).
- b) Analysieren Sie die worst-case Zeitkomplexität einer naiven Implementierung von UPGMA.
- c) Geben Sie eine Implementierung von UPGMA an, die eine bessere worst-case Laufzeit hat als die naive Implementierung.

Der Algorithmus UPGMA funktioniert folgendermaßen.

Eingabe: Distanzmatrix D

Initialisierung:

1. jedes Taxon i definiert ein Cluster C_i der Grösse $n_i = 1$
2. jedes Taxon bildet ein Blatt im Baum T .

Iteration, solange die Anzahl der Cluster ≥ 2 ist:

1. bestimme C_i und C_j mit $d(i, j)$ minimal
2. $C_k = C_i \cup C_j$ ist neues Cluster der Grösse $n_k = n_i + n_j$
3. erzeuge in T einen neuer Knoten mit Markierung $d(i, j)$, dessen Kinder die zu C_i und C_j korrespondierenden Knoten werden
4. füge zur Matrix d eine Reihe/Spalte für das Cluster C_k hinzu mit

$$d(k, \ell) = \frac{n_i \cdot d(i, \ell) + n_j \cdot d(j, \ell)}{n_i + n_j}$$

und lösche die Reihen/Spalten der Cluster C_i und C_j

Ausgabe: Baum T