Übungen Expressionsdatenanalyse I
Sommersemester 2007

Dr. Ivo Große

Institut für Informatik
Universität Halle

Blatt 1

Aufgabe 1.1

(a) Beweisen Sie die Identität aus der Vorlesung $\bar{d}_j = 2var(x_j)$.

(b) Zeigen Sie, dass der gewichtete Euklidische Abstand der Vektoren x mit Gewichten $w_j \propto 1/var(x_j)$ gleich dem ungewichteten Euklidischen Abstand der umskalierten Vektoren y mit $y_j = x_j/sqrtvar(x_j)$ ist. Welche Standardabweichung haben die Komponenten y_j?

(c) Zeigen Sie allgemeiner, dass der gewichtete Euklidische Abstand der Vektoren x mit beliebigen Gewichten w_j gleich dem ungewichteten Euklidischen Abstand der umskalierten Vektoren y mit $y_j = x_j/sqrtw_j$ ist.

Aufgabe 1.2

Finden Sie zwei Abstandsmatrizen, die nicht durch einen Baum repräsentiert werden können. Zeigen Sie explizit, dass Ihre Matrizen Abstandsmatrizen sind, d. h. dass sie die drei Abstandsaxiome erfüllen, und beweisen Sie, dass es für Ihre Matrizen keinen Baum geben kann.

Aufgabe 1.3

(a) Wenden Sie die drei agglomerativen hierarchischen Clusteralgorithmen (single linkage, average linkage, complete linkage) und die beiden in der Vorlesung behandelten divisiven hierarchischen Clusteralgorithmen auf die folgenden drei Matrizen per Hand an, und zeigen Sie in jedem Zwischenschritt die Teilbäume und die Kantenlängen. Zeigen oder widerlegen Sie für jede Matrix, ob sie eine Abstandsmatrix ist und ob sie ultrametrisch ist.

	A	B	C	D	E
A	0	2	8	8	8
B		0	8	8	8
C			0	4	4
D				0	2
E					0

Abgabe: 2. 5. 2007

	A	B	C	D	E
A	0	9	8	7	8
B		0	3	6	7
C			0	5	6
D				0	3
E					0

	A	B	C	D	E
A	0	2	5.2	8.8	13.4
B		0	3.2	6.8	11.4
C			0	3.6	8.2
D				0	4.6
E					0

(b) Finden Sie zwei Abstandsmatrizen, für die die drei hierarchischen Clusteralgorithmen (single linkage, average linkage, complete linkage) je drei verschiedene Bäume liefern.

(c) Beweisen oder widerlegen Sie die folgende Aussage: Falls die Abstandsmatrix ultrametrisch ist, liefern der single linkage, der average linkage und der complete linkage Algorithmus denselben ultrametrischen Baum.