Multivariate Statistik, Hausaufgabe 7

Henry Haustein

Aufgabe 1

(a) euklidische Distanz:

- $d(x_1, x_1) = 0$
- $d(x_1, x_2) = \sqrt{4^2 + 8^2} = \sqrt{80}$
- $d(x_1, x_3) = \sqrt{1^2 + 1^1} = \sqrt{2}$
- $d(x_2, x_2) = 0$
- $d(x_2, x_3) = \sqrt{3^2 + 7^2} = \sqrt{58}$
- $d(x_3, x_3) = 0$

Manhatten Distanz:

- $d(x_1, x_1) = 0$
- $d(x_1, x_2) = 4 + 8 = 12$
- $d(x_1, x_3) = 1 + 1 = 2$
- $d(x_2, x_2) = 0$
- $d(x_2, x_3) = 3 + 7 = 10$
- $d(x_3, x_3) = 0$

Für die Mahalanobisdistanz brauchen wir die Inverse der Varianz-Kovarianz-Matrix, die sich einfach mit R berechnen lässt. Es ergibt sich

$$S^{-1} = \begin{pmatrix} 14.25 & -6.75 \\ -6.75 & 3.25 \end{pmatrix}$$

und damit

- $d(x_1, x_1) = 0$
- $d(x_1, x_2) = \sqrt{4} = 2$
- $d(x_1, x_3) = \sqrt{4} = 2$
- $d(x_2, x_2) = 0$
- $d(x_2, x_3) = \sqrt{4} = 2$
- $d(x_3, x_3) = 0$

(b) Man sieht eigentlich recht schnell, dass zuerst x_1 und x_2 sowohl bei euklidischer als auch bei Manhatten Distanz zu einem Cluster zusammengefasst werden, danach wird x_3 diesem Cluster hinzugefügt. Bei der Mahalanobisdistanz haben alle Punkte den selben Abstand zueinander, man könnte sie also direkt in einem Cluster zusammenfassen.