Intelligent Data Analysis Summer 2017
C. Braune, T. Sabsch

# Exercise Sheet 10

**Exercise 37** Agglomerative Clustering

Let the following one-dimensional data set be given:

$$2, 5, 11, 12, 17, 21, 32.$$

Process this data set with hierarchical agglomerative clustering using

a) the centroid method,

b) the single linkage methode,

c) the complete linkage methode!

Draw a dendrogram for each case!

**Exercise 38** Method of Least Squares/Regression

Determine a best fit line $y = a + bx$ (regression line) for the data set already considered in exercise 10, that is, for

| $x$ | 0 | 1 | 1 | 2 | 3 | 3 | 4 | 5 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 0 | 1 | 2 | 3 | 2 | 3 | 4 | 4 | 6 | 5 |

a) using the covariance and the variances/standard deviations
(see the lecture slides, section on correlation coefficients)

b) using the method of least squares/the system of normal equations!

Draw a diagram of the data points and the regression line!

**Exercise 39** Method of Least Squares/Regression

Determine a best fit parabola $y = a + bx + cx^2$ (regression parabola) for the data set $(x, y) = ((0, 0), (2, 1), (3, 2), (4, 4))$ with the method of least squares and draw this parabola!

**Exercise 40** Multilinear Regression

Determine a best fit plane $z = a + bx + cy$ for the following data set with the method of least squares: $(x, y, z) = ((0, 1, 0), (0, 4, 2), (2, 0, 1), (3, 1, 2), (2, 3, 3), (4, 4, 4))$.