Intelligent Data Analysis Summer 2014
C. Braune

Exercise Sheet 9

Additional Exercise Expectation Maximization

In exercises 28 and 29 we processed the one-dimensional data set

$$1, 3, 4, 5, 8, 10, 11, 12$$

with fuzzy clustering and expectation maximization. In doing so we discovered a difference in the behavior of the membership degrees. In order to show that this difference does not result from the method, but from the distribution assumption, we apply the expectation maximization algorithm again. However, this time we do not assume a mixture of normal/Gaussian distributions, but a mixture of two Cauchy functions

$$f_{\text{cauchy}}(x; \mu; \sigma^2) = \frac{1}{\pi\sigma} \cdot \frac{1}{\frac{(x-\mu)^2}{\sigma^2} + 1}.$$

As in exercise 29 we assume that the prior probabilities are fixed to $\frac{1}{2}$ and the variances to 1. Likewise, the initial cluster centers are $\mu_1 = 1$ and $\mu_2 = 5$. Compute one expectation step and one maximization step!

Exercise 32 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 33 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 34 Multilinear Regression

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