Intelligent Data Analysis
R. Kruse, C. Braune

Summer 2012

Exercise Sheet 11

Exercise 35 Fuzzy Clustering

Consider the objective function of fuzzy clustering with a fuzzifier $w = 1$, that is,

$$J(\mathbf{X}, \mathbf{B}, \mathbf{U}) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}\, d^2(\beta_i, \vec{x}_j),$$

which is to be minimized under the constraint

$$\forall j \in \{1, \ldots, n\} : \quad \sum_{i=1}^{c} u_{ij} = 1.$$

Show that one obtains a hard/crisp assignment of the data points even if the membership degrees u_{ij} may come from the interval $[0, 1]$. That is, show that for the minimum of the objective function J it is $\forall i \in \{1, \ldots, c\} : \forall j \in \{1, \ldots, n\} : u_{ij} \in \{0, 1\}$.
(Hint: You may find it easier to consider the special case $c = 2$ (two clusters) and to examine the term for a single data point $\vec{x}_j$. Then generalize the result.)

Exercise 36 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 37 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 38 Logistic Regression

The following table shows the number of American intercontinental ballistic missiles (ICBMs) in the years from 1960 to 1969:

year, x	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
number, y	18	63	294	424	834	854	904	1054	1054	1054

Find a best fit curve for this data set using logistic regression ($Y = 1060$)! Draw the original data and sketch the curve $y = \frac{1060}{1+e^{a+bx}}$!

Exercise 39 Frequent Item Set Mining

Given the transactions from the table on the right determine all frequent item sets with a minimum support $s_{min} = 3$ using the apriori algorithm.
Which item sets are closed, which are maximal?

t_{ID}	items
1:	{ a,d,e }
2:	{ b,c,d }
3:	{ a,c,d }
4:	{ b,c,d,e }
5:	{ a,d }
6:	{ b,c,d }
7:	{ a,b,e }
8:	{ b,c,d,e }
9:	{ a,d,e }
10:	{ a,b,d }
11:	{ c,d }