Intelligent Data Analysis Summer 2014
C. Braune

Exercise Sheet 7

Exercise 24 Decision Trees: Attribute Selection Measures

Compute the information gain and the χ^2 measure for the following two contingency tables, which refer to two descriptive attributes A, B and one class attribute C!

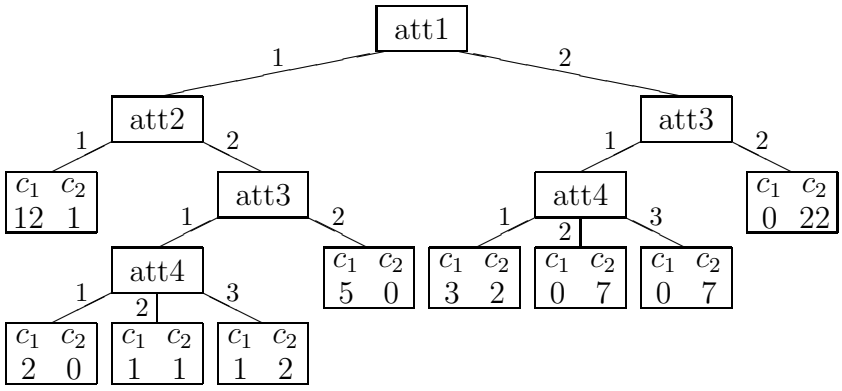
		A		
		a_1	a_2	a_3
C	c_1	9	4	3
	c_2	3	9	4
	c_3	4	3	9

		B		
		b_1	b_2	b_3
C	c_1	9	4	3
	c_2	6	6	4
	c_3	1	6	9

How may one describe the selection behavior of the two measures intuitively?
(Hint: Mind the first row and the last column of the two tables.)

Exercise 25 Decision Trees: Pruning

Prune the following decision tree using the approach of pessimistic pruning!
(parameter: 0.5 additional errors)



Exercise 26 c-Means Clustering

Consider the following two-dimensional data set:

x	1	6	8	3	2	2	6	6	7	7	8	8
y	5	2	1	5	4	6	1	8	3	6	3	7

Process this data set with c-means clustering with $c = 3$ (i.e., try to find 3 clusters)! Use the first three data tuples als initial positions for the cluster centers and observe the migration of the centers.

Exercise 27 c-Means Clustering

In exercises 16 and 17 on sheet 7 we considered a simple two-dimensional data set. Reconsider this data set, but assume that that no class information is available for the data points. That is, consider the following data set:

x	3	3	4	4	5	6	7	7	8	9	1	2	2	3	4	5	5	6	7	7
y	1	2	2	3	3	4	4	6	5	7	3	4	5	6	6	7	8	8	8	9

a) Which problem of c-means clustering becomes obvious when this data set is processed with $c = 2$ (i.e., if one tries to find two clusters)?
Hint: What is the desired result? What is produced by c-means clustering?
(You need not compute the exact result of the algorithm, a qualitative description suffices. Compare the result to a naive Bayes classifier.)

b) How could one try to cope with this problem?
Hint: Recall what distinguishes a full and a naive Bayes classifier.

Additional Exercise Lagrange Theory

Determine the minimum of the function $f(x, y) = xy^2 + x + 2y$ under the constraints $xy = 1$ and $x > 0$ with the help of the method of Lagrange multipliers!