Intelligent Data Analysis Summer 2013
C. Braune

Exercise Sheet 5

Exercise 17 Naive Bayes Classifier

Determine a naive Bayes classifier from the data set shown below! Assume that x and y are continuous attributes and that they are normally distributed (per class).

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
Class	a	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b

Classify the following two cases with the obtained classifier:

x	8	3
y	7	4

Is the result plausible? (In order to answer this question it may be helpful to draw the data set and the two sample cases to classify in an appropriate coordinate system.)

Exercise 18 Full Bayes Classifier

Use the dataset given in exercise 17 to determine a full instead of a naive Bayes classifier and classify the two test cases with this classifier! Again assume that x and y are continuous attributes and that they are normally distributed (per class). How does the result differ from the result obtained in exercise 17? Is this new result plausible?

Exercise 19 Visualizing Bayes Classifiers

Visualize the two classifiers obtained in exercises 17 and 18 by drawing the dataset, the center/mean vectors, and the $2\hat{\sigma}$ ellipses of the two-dimensional normal distributions in an x-y coordinate system. (Hint: Compute an appropriate transformation matrix with Cholesky decomposition or eigenvalue decomposition. You may also check your result with the Bayes classifier and visualization programs available on the lecture page.) Check the classification results obtained in exercises 17 and 18! Which disadvantage/restriction of the naive Bayes classifier becomes obvious with this visualization?

Exercise 20 Bayes Classifiers: Benefit of Attributes

Right or wrong? The number of cases a

a) naive Bayes classifier
b) full Bayes classifier

misclassifies of the training data (that is, the data set they are induced from) can only decrease if more descriptive attributes are used.