

Exercise Sheet 7

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 as initial positions for the cluster centers and observe the migration of the centers.

Exercise 27 c -Means Clustering

In exercises 17 and 18 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

- 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.)
- 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!

Exercise 28 Fuzzy Clustering

Consider the one-dimensional data set

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

We want to process this data set with fuzzy c -means clustering with $c = 2$ (two clusters) and a fuzzifier of $w = 2$. Assume that the cluster centers are initialized to 1 and 5. Execute one step of alternating optimization as it is used for fuzzy clustering, i.e.:

- a) Compute the membership degrees of the data points for the initial cluster centers!
- b) Compute new cluster centers from the membership degrees computed in this way!

Exercise 29 Expectation Maximization

Consider again the one-dimensional data set used in exercise 28, which we want to process in this exercise with the expectation maximization algorithm to estimate the parameters of a mixture of two normal/Gaussian distributions. Let the prior probabilities of the two clusters be fixed to $\theta_i = \frac{1}{2}$ and the variances to $\sigma_i^2 = 1$, $i = 1, 2$. (That is, only the expected values of the normal distributions — the cluster centers — are to be adapted.) Use the same values for the initial cluster centers as in exercise 28, that is, 1 and 5. Compute one expectation step and one maximization step, i.e.:

- a) Compute the posterior probabilities of the data points for the initial cluster centers!
- b) Estimate new cluster centers from the data point weights computed in this way!