Prof. Dr. Hauke Schramm
Department of Computer Science
Kiel University

Pattern Recognition

Exercises

Practice Sheet 4

Remark

Voluntary exercises – no submission required!

Exercise L-4.1 (Normal distribution, mixture density)

In this exercise, you will learn how to model unknown joint probability distributions $p(x_1, x_2)$ in order to describe a given dataset. To this end, we will use Normal distributions with independent or dependent components.

Load the datasets a.mat and x.mat or, alternatively, a.txt and x.txt from the OLAT material folder. Each dataset contains N two-dimensional data points (a.mat: N = 10.000, x.mat: N = 20.000) which are stored in two columns. Example: The first ten data points of dataset x.mat can be accessed by `x(1:10, :)` and the 5th data point by `x(5, :)`. Study the points using the plot command: `plot(.., ".")`.

Data set a.mat

Start modeling the data set a.mat with a Gaussian distribution with independent components. Visualize the data points and superimpose them with the contour image of the Gaussian using the Octave function `contour()` (see the lecture for details). Additionally generate a second figure with the 3D mesh of the estimated Gaussian using the function `mesh()`. This mesh represents the joint probability $p(x_1, x_2)$ (we may also write $p(X)$ with $X=[x_1, x_2]^t$) in z-direction over the (x_1, x_2) plane. Combine both images into a single figure using the `subplot()` function. Does the distribution represent the data sufficiently good?

Repeat the described procedure but this time using a Gaussian distribution with *dependent* components. Do you achieve a better modeling of the data?

Combine all 4 images into a single overview figure with correct labeling of the axes (commands: `xlabel, ylabel`) and reasonable title (command: `title`).

Prof. Dr. Hauke Schramm
Department of Computer Science
Kiel University

Institut für Informatik
Christian-Albrechts-Universität zu Kiel

C A U
Christian-Albrechts-Universität zu Kiel

Data set x.mat

Model the data set **x.mat** with a Gaussian with dependent / independent components. Generate overview images showing the point distribution with overlay contours and corresponding meshes and discuss the quality of the model. Are you satisfied with your modeling?

Model this distribution with a mixture density and use k-means clustering (cf. lecture) to estimate the required parameters. How do you reasonably determine the number of required splits and iterations?

Study the resulting contour images and discuss if the modeling has been improved.