Homework Assignment 3

Name(s) Matr.No(s).

The analytical part of your homework as well as a short simulation protocol for the MATLAB part has to be delivered as hard copy at the beginning of the lecture on **2011/4/1**. Use a printed version of this assignment document as the title page(s) and fill in your name(s) and matriculation number(s). You may work in pairs (only one copy needs to be delivered per pair).

The MATLAB programs of the homework have to be submitted via e-mail to `hw2.spsc@tugraz.at` no later than **2011/4/1**. The subject of the e-mail consists of your last name(s) and the assignment number (e.g., "Kullback, Leibler: Assignment 3"). Make sure your programs contain a lot of helpful comments.

Problem 3.1 (12 Points)

Let $\mathbf{X}$ be a two-dimensional random vector with independent components, each with a uniform distribution on the interval $[0, 1)$. Consider the two-dimensional random vector $\mathbf{Y} = B\mathbf{X}$, where

$$B = \begin{bmatrix} 1 & 0.2 \\ 1 & -0.2 \end{bmatrix}.$$

(a) Compute $\mu_{\mathbf{Y}} = E\{\mathbf{Y}\}$ and $\mathbf{C}_{\mathbf{Y}} = E\left\{(\mathbf{Y} - E\{\mathbf{Y}\})(\mathbf{Y} - E\{\mathbf{Y}\})^{\mathsf{T}}\right\}$ analytically. Hint: express these expectations in terms of expectations with respect to $\mathbf{X}$.

(b) Compute the differential entropy $h(\mathbf{Y})$ analytically.

(c) For this and the following tasks, MATLAB scripts (or functions) need to be written. Generate 100 000 realizations of the random vector $\mathbf{Y}$. We refer to these realizations as the training data. Make a scatter plot of the training data.

(d) Given the training data, estimate the parameters of a multivariate Gaussian density model (i.e, its mean vector and covariance matrix) in the sense of a *maximum likelihood*. Compare your result with that obtained in **(a)**.

(e) Implement the EM algorithm to optimize the parameters of a Gaussian mixture model (GMM) for a vector random variable (see Example 4.4 in our lecture notes). For simplicity, the algorithm stops after a given (sufficiently high) number of iterations. Using your implementation of the EM algorithm and the training data, train a GMM with *two* mixture components (*full* covariance matrices).

(f) Train a GMM with *two* mixture components with *diagonal* covariance matrices. (Do not estimate nor update off-diagonal entries of the covariance matrices, but simply set them to zero.)

(g) Train a GMM with *four* mixture components with *diagonal* covariance matrices.

(h) Visualize the model densities obtained in the previous tasks **(d)–(g)** (e.g. by evaluating those densities at grid points and using `pcolor()`). Compare the obtained plots with the scatter plot of the training data.

(i) Download `density_model_quality_measure.m` from our course web page. For the given training data and the model densities obtained in the previous tasks **(d)–(g)**, compare their approximation qualities using the downloaded function. Explain what this measure approximates. Are we interested in minimizing or in maximizing this measure? Also, compare with $h(\mathbf{Y})$ obtained in **(b)**.