Exercises for
Pattern Analysis
Sebastian Käppler
Assignment 7, 16.06.2015

General Information:

Lecture (3 SWS):	Thu 14.15 – 15.45 (H16) and Tue 12.15 – 13.45 (H16)
Exercises (1 SWS):	Mo 12.15 – 13.45 (02.134-113) and Tue 12.15 – 13.45 (E1.12)
Certificate:	Oral exam at the end of the semester
Contact:	sebastian.kaeppler@fau.de

Probability Density Estimation - Part II

Exercise 1 The mean shift algorithm can be used to determine a local maximum (or a saddle point) of a probability density function $p(\boldsymbol{x})$. In particular, it is feasible to determine the *mode* of the density. For a mathematical formulation of the mean shift algorithm, the Parzen window approach is used to model $p(\boldsymbol{x})$. The mean shift iterations are equivalent to a gradient ascent for $p(\boldsymbol{x})$.

(a) Outline the main steps of the mean shift algorithm.

(b) Derive the mean shift vector for the following kernels:

- Epanechnikov kernel
- Gaussian kernel

Exercise 2 Let $\mathcal{S} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a set of $N = 8$ samples defined as:

$$\mathcal{S} = \left\{ \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}, \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}, \begin{pmatrix} 0.2 \\ 0.25 \end{pmatrix}, \begin{pmatrix} 0.3 \\ 0.2 \end{pmatrix}, \begin{pmatrix} 0.5 \\ 0.7 \end{pmatrix}, \begin{pmatrix} 0.7 \\ 0.8 \end{pmatrix}, \begin{pmatrix} 0.8 \\ 0.9 \end{pmatrix}, \begin{pmatrix} 0.9 \\ 0.8 \end{pmatrix} \right\}$$

(a) Draw the samples in the 2-dimensional feature space.

(b) Perform one mean shift iteration and draw the corresponding mean shift vectors using the following starting points:

- $\boldsymbol{x}^0 = \begin{pmatrix} 0 & 0 \end{pmatrix}^\top$
- $\boldsymbol{x}^0 = \begin{pmatrix} 1 & 1 \end{pmatrix}^\top$

Use the Epanechnikov kernel with kernel width $\lambda = 0.25$.

(c) Sketch the mean shift vectors if the mean shift iterations are performed until convergence.

(d) Explain how the mean shift algorithm can be used for an automatic clustering. How do you determine the number of clusters? Compare mean shift clustering to hard- and soft-clustering.

Exercise 3 **Python exercise** Download the updated classification toolbox from the exercise website. Support for 2-D histogram and density estimation has been added. Your task is to implement histogram estimation in *HistogramDensityEstimation.py* and Kernel Density Estimation in *KernelDensityEstimation.py*.

(a) For Histogram estimation, *coordinateSystem.getLimits()* will give you the bounds of the coordinate system for histogram estimation (i.e. min/max values of the histogram), while *binsPerUnitX/Y* determine the number of bins.

(b) For Kernel density estimation, use a Gaussian Kernel and output log probabilities for easier visualization. Don't forget to normalize the result.