Machine Intelligence II **Exercise Sheet 3**
SS2012, Prof. Obermayer due: 03.05.2012

Probability Density Estimation

In the exercise we will estimate a probability density with a kernel estimator and find the appropriate hyperparameter by minimizing the log-likelihood. Use Matlab or Python for this programming exercise. ***Data:*** Download the image file `testimg.jpg` from the ISIS platform.

3.1 Data

(3 points)

- Load the image `testimg.jpg`. The intensities are given by one byte per pixel, i.e. as gray values represented by type `unsigned char` (or `uint8`, or similar). Thus $I \in [0, 255]^{m \times n}$, where I is the image and m, n are the height and width of the image.

  Hint: For some calculations it could be necessary to convert the values to some floating point type (e.g. `double`)
- Consider the pixel values as i.i.d. samples. That is, the values are assumed to be drawn all from the same probability distribution and not influenced by the other values (which is not true in this case, but this doesn't matter for the moment).
- Add Gaussian noise with standard deviation σ_N. For the specific values of σ_N see below.

  Hint: The standard deviation σ_N is specified for a normalized pixel range, that is, $I \in [0, 1]^{m \times n}$. If you work with a value range of a byte ($I \in [0, 255]^{m \times n}$), you have to scale σ_N appropriately.

3.2 Kernel Density Estimation

(3 points)

Estimate the probability density of intensities with a rectangular or Gaussian kernel ("gliding window") characterized by a width parameter h (e.g. window width or standard deviation). Use a random subset of the samples of size P to estimate the density and plot the estimates $\hat{p}$ resulting for different widths. For the specific values of P see below.

3.3 Validation

(4 points)

Calculate the negative log-likelihood of your estimator for each kernel width h using all pixels, which were not used for density estimation. Plot the results against h. Do this for each combination of two values for the noise standard deviation (e.g. $\sigma_N \in \{0.05, 0.1\}$) and two values for the sample set size (e.g. $P \in \{100, 500\}$). Which kernel width h yields the minimal negative log-likelihood hence the minimal generalization error, for each combination?

Hint: Feel free to use different parameters if you find no local minimum in the specified range.

Total points: 10