Machine Learning WS2014/15 (Unsupervised Learning)

Lecturer: Prof. Dr. Bethge

3. Continuous distributions and generative models

Task 1:

Show that the maximum likelihood estimates of the parameters of a multivariate Normal distribution $\mathcal{N}(\mathbf{x}|\mathbf{m}, C)$ are given by the sample mean and the sample covariance.

Task 2:

Consider the following generative model:

$$\underbrace{\mathbf{u} \sim \prod_{k=1}^{d} U[-1, 1]}_{\text{probability of u}}, \qquad \mathbf{x} = A\mathbf{u}$$

where $U[a, b]$ denotes the uniform distribution between a and b.

1. Compute the differential entropies $h[p(\mathbf{u})]$ and $h[p(\mathbf{x})]$ in case when
   (i) A is invertible
   (ii) A is orthogonal.

2. Estimate numerically the cross-entropy of the maximum likelihood Gaussian model.

3. In the following always ssume $d = 2$ and $A = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. Compute the probability density function $p(x_k)$ and compare it to $p(u_k)$.

4. Compute and compare the variance of the components u_k to the variance of the components of x_k.

5. Sample from the distribution $\hat{p}(\mathbf{x}) = p(x_1)p(x_2)$ and show the data points in a scatter plot. Then sample from the true distribution $p(\mathbf{x})$ and superimpose the data points in the same scatter plot but with a different color.

6. Estimate numerically the cross-entropy of $\hat{p}(\mathbf{x}) = p(x_1)p(x_2)$ and compare it to the Gaussian cross-entropy and to the true entropy.

7. Modify the generative model by additionally multiplying $\mathbf{x}$ with an independent scalar variable z drawn from a Gamma distribution with parameters $a = 2$ and $b = 1/2$ such that $\mathbf{x} = z \cdot A\mathbf{u}$. Still assume $d = 2$ and $A = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. Estimate numerically the cross-entropy of $\hat{p}(\mathbf{x}) = \mathcal{L}(x_1|0, \sigma)\mathcal{L}(x_2|0, \sigma)$ where $\mathcal{L}(x_k|0, \sigma)$ denotes the Laplacian distribution with scale σ. The scale σ should be chosen to maximize the likelihood (i.e. minimize the cross-entropy).

   [Optional extra task: Compute the true density $p(\mathbf{x})$ and use it to numerically estimate the true differential entropy of the data]