UML computer project 2

- Handed out: April 14 (Thu)
- Hand in: May 2 (Mo) noon the latest,
  by email to michael.gutmann@helsinki.fi
- You can do the exercises in pairs. Submit in that case only one report. Please write in the report the names and the student numbers of both of you.
- Submit your report as a single pdf containing figures and **discussion**. In the report, explain what you are doing and why you are doing it. Don't put figures in the report without explaining the result they show. The report should be enjoyable to read; remember that the grading will be based on the report.
- Submit the source code as well (separate attachment, not as appendix in the report). The code needs to be such that running it for every exercise will produce the figures in the report.
- Each exercise in the project gives you points. They will, at the end, determine your grade for the computer project. Each exercise in all the assignments gives you an equal amount of points.

Exercise 1: Basics of ICA

Figure 1 shows data which follows a uniform distribution of mean zero and variance one, as well as that data after the linear transformations A_1 and A_2,

$$A_1 = \begin{pmatrix} 0.4483 & -1.6730 \\ 2.1907 & -1.4836 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & -1.7321 \\ 1.7321 & -2.0 \end{pmatrix}. \qquad (1)$$

The data contains 5000 data points. In the following, we denote by $\mathbf{s}$ the random vector corresponding to the original data (left plot in the figure) and by $\mathbf{x}_1 = A_1\mathbf{s}$ and $\mathbf{x}_2 = A_2\mathbf{s}$ the random vectors after the linear transformation.

1. For a random vector $\mathbf{n} \in \mathbb{R}^2$ which follows a standard Gaussian distribution (mean zero, identity covariance matrix), make scatter plots showing $\mathbf{y}_1 = A_1\mathbf{n}$ and $\mathbf{y}_2 = A_2\mathbf{n}$ (use again 5000 data points). Describe the data after the linear transformations (distribution, mean, covariance matrix).

2. Whiten the four data sets corresponding to $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{y}_1$ and $\mathbf{y}_2$, and show scatter plots of the whitened data. Compare the whitened data to the original data before the linear transformations. How do the whitening matrices relate to the matrices A_i? (p. 34, section 4.5 in the lecture notes is about whitening. After whitening, the uniform data should look like the data in Fig.6.7 in the lecture notes.)

3. For all four data sets: project the data onto the unit vector $\mathbf{w}$ which forms an angle α, $0 \leq \alpha \leq \pi$, with the x-axis, and compute the kurtosis of the projection in function of α (You should get four curves, each like the curve shown in Fig.7.11). For which angles is the absolute value of the kurtosis maximized? Explain the qualitative difference in the curves for uniform data and Gaussian data.

4. How can you obtain an estimate for the A_i using the whitening matrices and the optimal projection vectors (those which maximize the absolute value of the kurtosis)? What are the estimates you obtain for A_1 and A_2?

5. Explain why finding an estimate for A_1 and A_2 is possible for the uniform data but not for the Gaussian data.

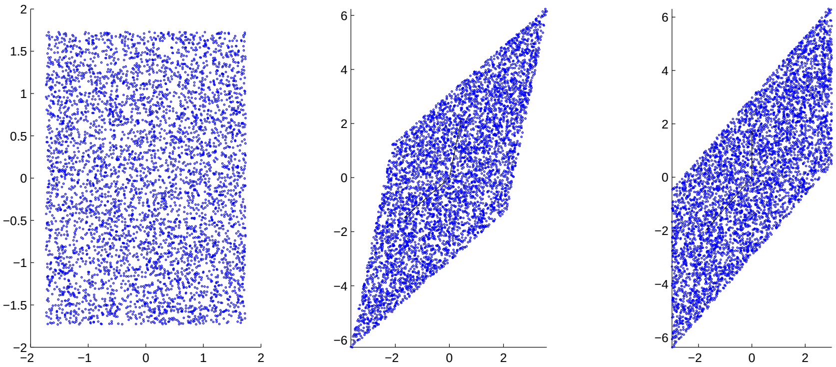

Figure 1: For exercise 1. Left: data points that follow a uniform distribution of mean zero and variance one. Middle: The data points after transformation with A_1. Right: The data points after transformation with A_2. In each of the three figures, 5000 data points are shown.

Exercise 2: Kurtosis based ICA

Let $\mathbf{s} = (s_1, \dots, s_{32})^T$ be a random vector which consists of 32 independent random variables, all of which follow a Laplacian distribution of mean zero and variance one ($s = 1/\sqrt{2}*\text{sign}(u)*\log(1-2|u|)$ follows such a distribution if u is a uniform random variable on $[-0.5\ 0.5)$).

1. Generate 10000 samples of $\mathbf{s}$. Compute an estimate of the probability density function (pdf) of y_m

$$y_m = \frac{\tilde{y}_m}{\sqrt{\text{Var}(\tilde{y})}} \qquad\qquad \tilde{y}_m = \sum_{i=1}^{m} s_i \tag{2}$$

   for $m = 1, 2, 4, 8, 16, 32$. The variable y_m is the sum of the m first s_i, normalized to unit variance. Show the logarithm of the six pdfs and compare the curves to the log-distribution of a Gaussian. (Hint: the estimates for the pdfs can be obtained via a histogram.)

2. Compute the kurtosis of y_m in function of $m \in \{1, \dots, 32\}$. Explain the behavior of the curve. Relate it also to the log-pdfs which you obtained in the previous question.

3. Implement the kurtosis-based ICA algorithm in section 7.4.3 in the lecture notes. Test it on one of the data sets of Exercise 1. The algorithm should return the demixing matrix as well as the values of the objective function during the optimization. Take care to implement a proper stop criterion in the optimization.

4. Let $\mathbf{y} = (y_1, \dots, y_{32})^T$. The transformation $\mathbf{s} \to \mathbf{y}$ can be written as linear transformation A. What is the formula to obtain A ?

5. Use ICA to get an estimate $\hat{A}$ of A from the observations of $\mathbf{y}$ and $\mathbf{s}$ alone. What is the average squared error $1/(32^2)\sum_{ij}(A_{ij} - \hat{A}_{ij})^2$? How large is the error if you use 20000 samples instead of 10000? (Note: to compute the error, you must take into account that ICA delivers only results up to a permutation matrix and sign flips of the columns of $\hat{A}$, see Section 6.3.3 in the lecture notes)

Exercise 3: Separating mixtures of images

In this exercise, we work with the ICA model $\mathbf{x} = A\mathbf{s}$ for $\mathbf{s} \in \mathbb{R}^6$ and $\mathbf{x} \in \mathbb{R}^6$: I took 6 different images and mixed them randomly together. The size of each image was $300px \times 300px$, and for the mixing, I stacked the columns of each image on each other to form a long 90000 dimensional (row) vector. The file `mixed_images.txt` contains the resulting mixture. The data matrix X has dimension 6×90000. The goal of this exercise is to recover the original images from the mixture.

1. Load the data and visualize each mixture as an image of size $300px \times 300px$. In ICA, the observations (data points) are assumed identically distributed (see likelihood based ICA, section 8.1). How well do we respect this assumption in this exercise? Explain what the random vector $\mathbf{s}$ is representing.

2. Visualize the whitened data. Compared to the mixture, can you already better guess what the original images are like?

3. Implement the ICA algorithm below. It is a modified version of the one in Table 8.1 in the lecture notes. Test it on some data from Exercise 1 (show also the values of F during the optimization and the values of γ_i after the optimization).
   ICA algorithm:

   (a) Whiten the data (the corresponding random vector is denoted by $\mathbf{z} \in \mathbb{R}^n$ below).

   (b) Initialization: random for the $n \times n$ matrix B, $\gamma_i = 0$ $(i = 1 \ldots n)$, $\mu_g = 0.8$, $\mu = 0.2$ (this are just possible values which worked fine for me).

   (c) Compute $\mathbf{y} = B\mathbf{z}$.

   (d) Update γ_i:

   $$\gamma_i \leftarrow (1 - \mu_g)\gamma_i + \mu_g \, \mathrm{E}(-\tanh(y_i)y_i + (1 - \tanh(y_i))^2)$$

   If $\gamma_i > 0$ define $g_i(u) = -2\tanh(u)$, else as $g_i(u) = \tanh(u) - u$

   (e) Compute the objective $F = -\sum_i \gamma_i \, \mathrm{E}(\log\cosh(y_i))$

(f) Update B by

$$B \leftarrow B + \mu(I + \mathrm{E}(\mathbf{g}(\mathbf{y})\mathbf{y}^T))B$$

where $\mathbf{g}(\mathbf{y}) = (g_1(y_1), \ldots, g_n(y_n))^T$ and I is the identity matrix.

(g) Orthonormalize B.

(h) Check convergence: If the change in F is smaller than some small threshold go back to step (c). Else return B and all the values of F.

4. Apply the ICA algorithm to demix the images. Show the resulting images and comment on the quality of the demixing (Note: you may need to flip the signs of the obtained images). Look at the values of the learned γ_i: what does this imply for the six different distributions of the sources s_i ?