Machine Intelligence 2 **Exercise Sheet 6**
SoSe 2022, Obermayer/Kashef/Strömsdörfer due: 2022-06-16 23:55

Maximizing nongaussianity

Exercise T6.1: Solving the ICA problem by maximizing nongaussianity (tutorial)

(a) What are the ambiguities and the limitations of the solutions found by ICA?

(b) What role does *whitening/sphering* play in the context of ICA?

(c) Why are Gaussians bad for ICA?

(d) How do we find independent components by maximizing nongaussianity?

(e) What measures do we have for nongaussianity and how do we use each for solving the ICA problem?

Exercise H6.1: Kurtosis of Toy Data (homework, 6 points)

The file `distrib.mat` contains three toy datasets (`uniform`, `normal`, `laplacian`)[1]. Each is made up of $10,000$ samples with 2 sources (i.e. $N = 2, p = 10,000$).
You are asked to do the following for each dataset:

(a) Apply the following mixing matrix $\underline{\mathbf{A}}$ to the original sources $\underline{\mathbf{s}}$:

$$\begin{aligned}\underline{\mathbf{A}} &= \begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix} \\ \underline{\mathbf{x}} &= \underline{\mathbf{A}}\,\underline{\mathbf{s}}.\end{aligned}$$

(b) Center the mixtures $\underline{\mathbf{x}}$ to zero mean.

(c) Decorrelate the mixtures from (b) by applying principal component analysis (PCA) on them and project them onto the PCs.

(d) Scale the decorrelated mixtures from (c) to unit variance in each PC direction. The mixtures are now *whitened* (*sphered*).

(e) Rotate the whitened mixtures by different angles θ

$$\begin{aligned}\underline{\mathbf{x}}_\theta &= \underline{\mathbf{R}}_\theta\,\underline{\mathbf{x}} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \underline{\mathbf{x}} \\ \theta &= 0, \frac{\pi}{50}, \ldots, 2\pi\end{aligned}$$

and calculate the (excess) kurtosis[2] empirically for each dimension in $\underline{\mathbf{x}}$:

$$\mathrm{kurt}(x_\theta) = \left\langle x_\theta^4 \right\rangle - 3 \underbrace{\left\langle x_\theta^2 \right\rangle}_{=1}^2.$$

[1] Python users can load the content of `.mat` files into a `dict` using the function `loadmat` from `scipy.io`

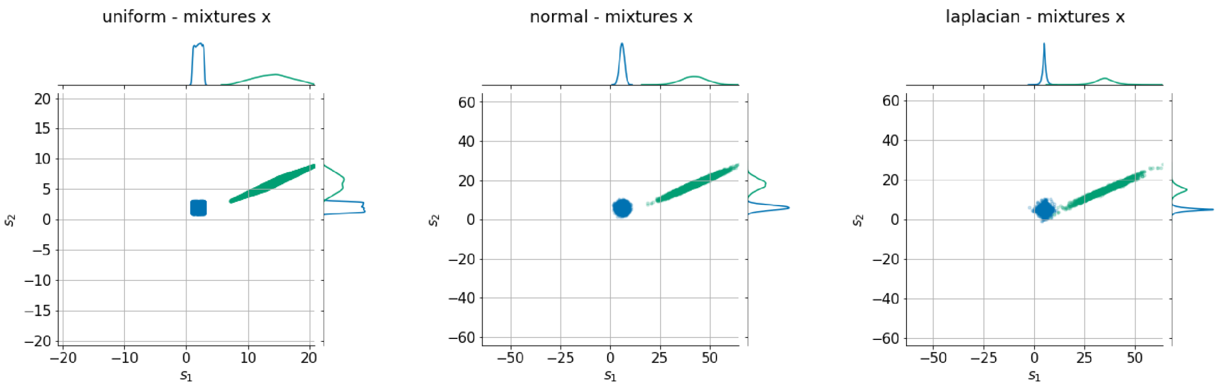
[2] Here and in the lecture notes the so-called *excess* Kurtosis is used which yields a value of 0 for normally distributed random variables. Additionally, this definition does not explicitly normalize by the standard deviation, because the standard deviation of each dimension is 1 after whitening.

(f) Find the minimum and maximum kurtosis value for the first dimension and rotate the data accordingly.

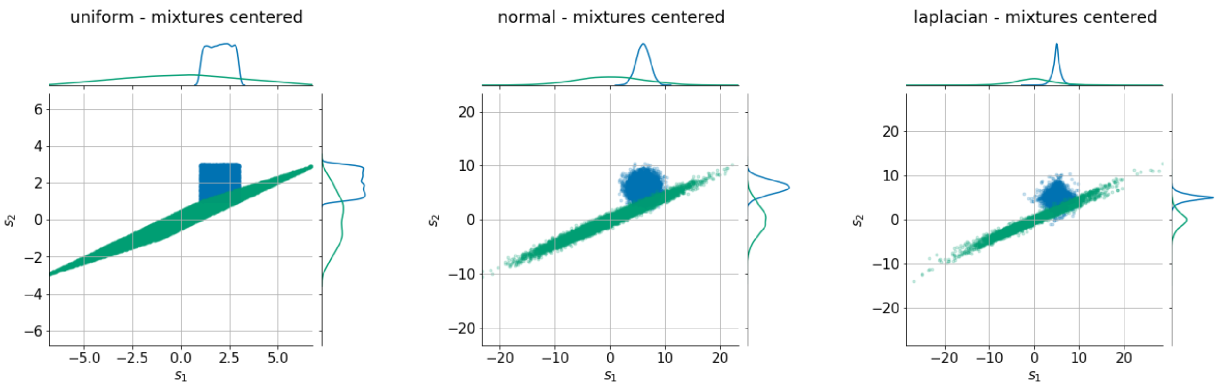
- Plot the original dataset (sources) and the mixtures after the steps (a), (b), (c), (d), and (f) as a scatter plot and display the respective marginal histograms.
- For step (e) plot the kurtosis value $\mathrm{kurt}(x_\theta)$ of each dimension in $\underline{\mathbf{x}}$ as a function of the rotation angle θ for each dimension.
- Compare the histograms after rotation by θ_{min} and θ_{max} for the different distributions.

Solution

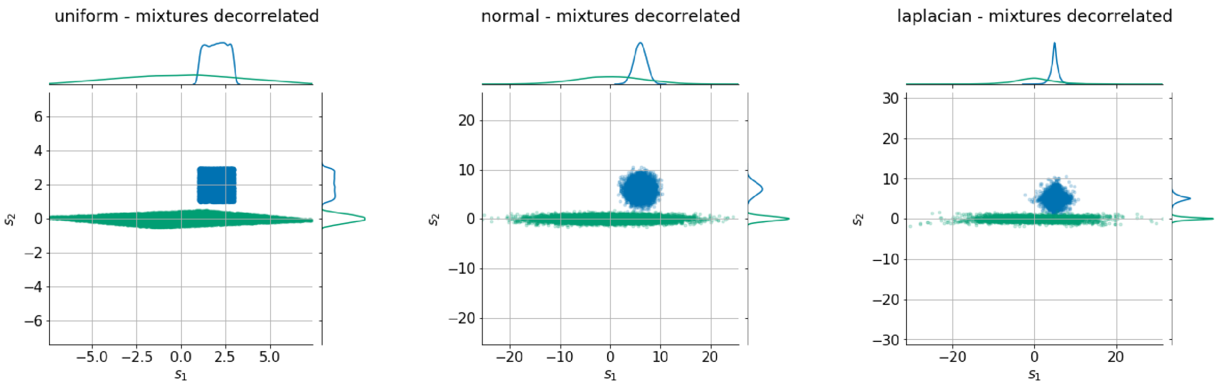
(a)



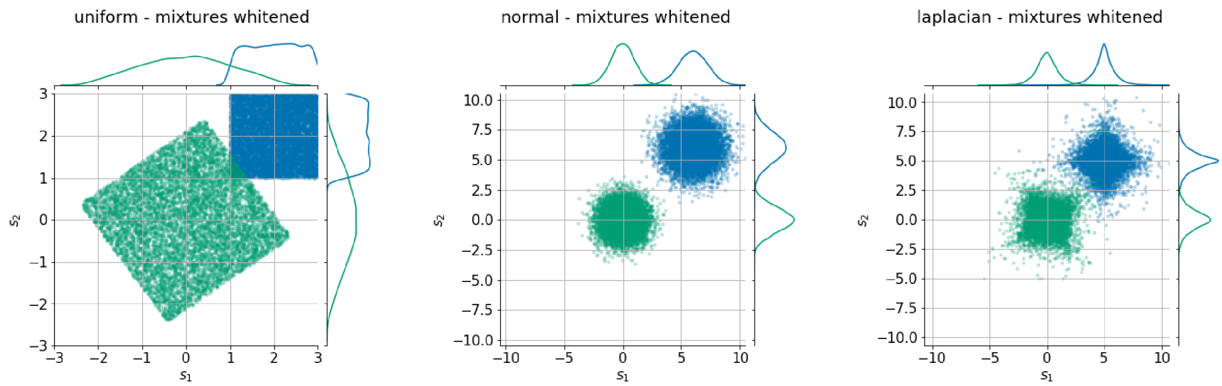
(b)


uniform - mixtures centered
normal - mixtures centered
laplacian - mixtures centered
s_1
s_2

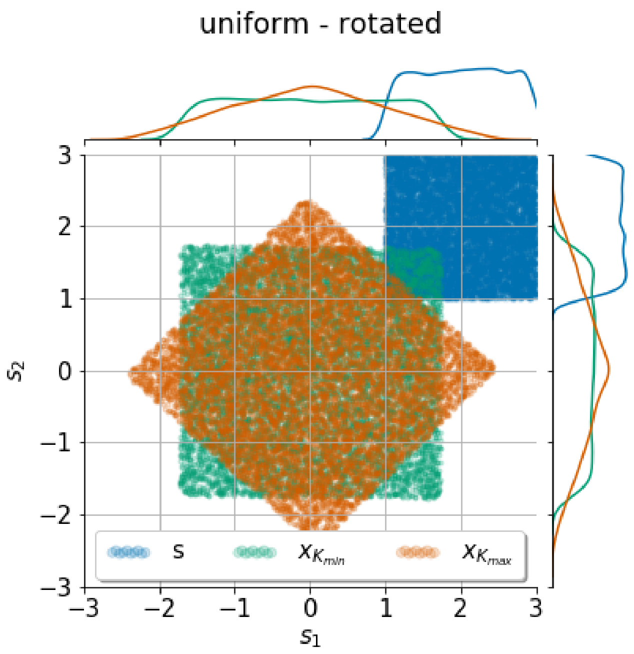
(c)

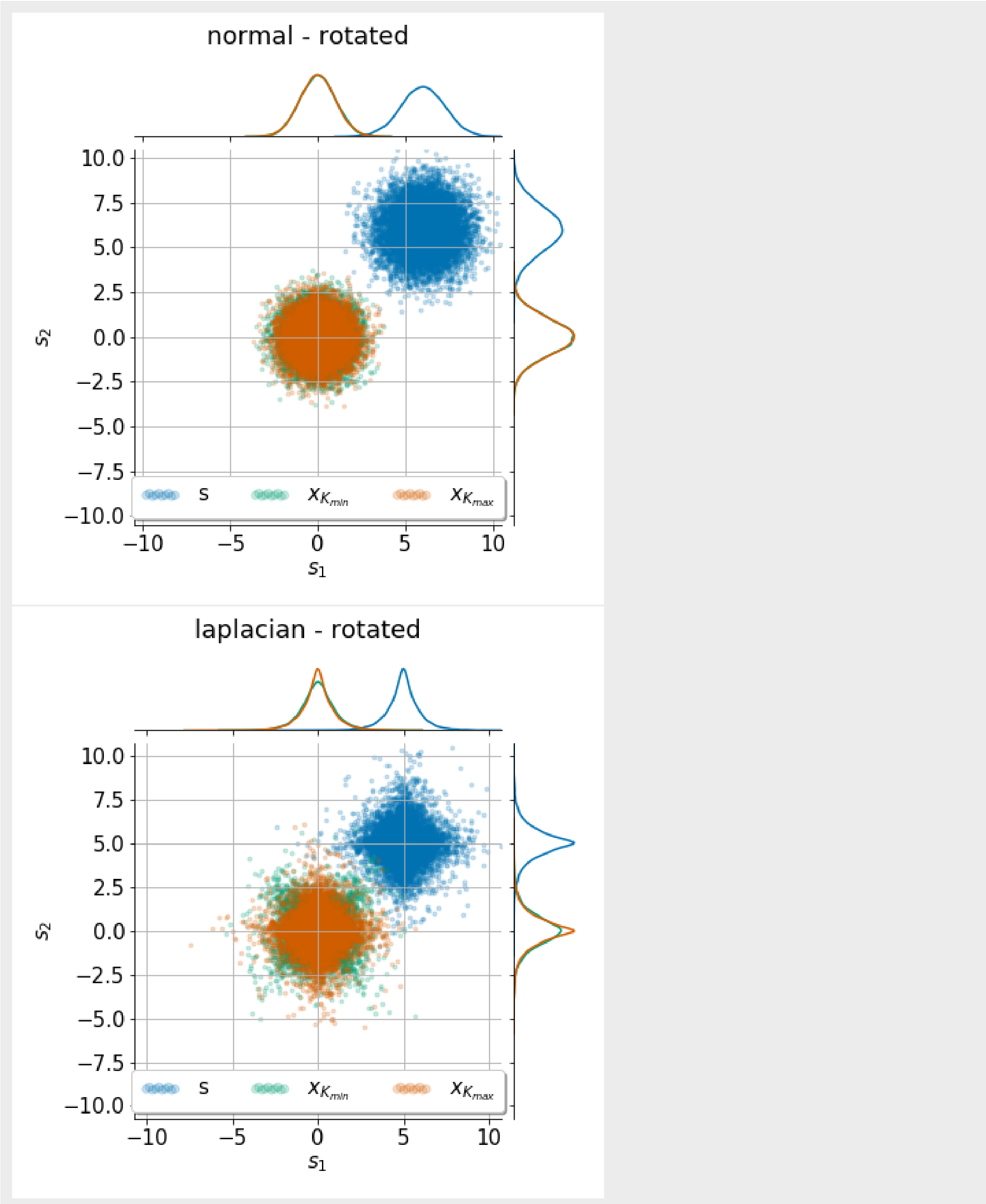

uniform - mixtures decorrelated
normal - mixtures decorrelated
laplacian - mixtures decorrelated
s_1
s_2

(d)


uniform - mixtures whitened
normal - mixtures whitened
laplacian - mixtures whitened
s_1
s_2

(f)


uniform - rotated
s_1
s_2
s
$x_{K_{min}}$
$x_{K_{max}}$



Exercise H6.2: Negentropy is scale-invariant **(homework, 4 points)**

The differential entropy of an N-dimensional random vector $\underline{X}$ with probability density $p(\underline{\mathbf{x}})$ is defined as

$$H(\underline{X}) = -\int_{\mathbb{R}^N} p(\underline{\mathbf{x}}) \log\, p(\underline{\mathbf{x}}) d\underline{\mathbf{x}}$$

The negentropy is defined as

$$J(\underline{X}) = H(\underline{X}_{Gauss}) - H(\underline{X})$$

where $\underline{X}_{Gauss}$ is an N-dimensional multivariate Gaussian random vector with the same covariance matrix as $\underline{X}$.

Show that the negentropy is invariant w.r.t. to an invertible ($\det \underline{\mathbf{A}} \neq 0$) linear transformations $\underline{\mathbf{y}} = \underline{\mathbf{A}}\,\underline{\mathbf{x}}$, i.e.

$$J(\underline{\mathbf{A}}\,\underline{X}) = J(\underline{X})$$

from which it follows that the negentropy is scale-invariant.

Use that the differential entropy of a multivariate N-dimensional Gaussian random vector $\underline{X}$ with covariance matrix $\underline{\mathbf{\Sigma}}$ has the form

$$H(\underline{X}_{Gauss}) = \frac{1}{2}\,\log\,|\det\,\underline{\mathbf{\Sigma}}| + \frac{N}{2}\,(1 + \log\,2\pi)$$

Remark: Differential entropy itself is not scale-invariant.

Solution

The Covariance matrix of the random vector $\underline{\mathbf{y}} = \underline{\mathbf{A}}\,\underline{\mathbf{x}}$ is given as

$$\underline{\mathbf{\Sigma}}_Y = \mathbb{E}[\,\underline{\mathbf{A}}(\underline{\mathbf{x}} - \underline{\mu})(\underline{\mathbf{x}} - \underline{\mu})^\top \underline{\mathbf{A}}^\top] = \underline{\mathbf{A}}\,\underline{\mathbf{\Sigma}}\,\underline{\mathbf{A}}^\top$$

Let $\underline{Y}_G$ be a gaussian random vector with the same covariance matrix $\underline{\mathbf{\Sigma}}_Y$. Because $\underline{\mathbf{\Sigma}}$ is symmetric, therefore $|\underline{\mathbf{\Sigma}}| = |\underline{\mathbf{\Sigma}}^\top|$

We first find an expression for $H(Y)$:

$$\begin{aligned}
H(\underline{Y}) &= -\int p(\underline{\mathbf{y}}) \log\, p(\underline{\mathbf{y}}) d\underline{\mathbf{y}} \\
&= -\int p(\underline{\mathbf{x}}) \frac{1}{\left|\frac{d\underline{\mathbf{y}}}{d\underline{\mathbf{x}}}\right|} \log \left(p(\underline{\mathbf{x}}) \frac{1}{\left|\frac{d\underline{\mathbf{y}}}{d\underline{\mathbf{x}}}\right|} \right) d\underline{\mathbf{y}} \\
&= -\int p(\underline{\mathbf{x}}) \frac{1}{|\det \underline{\mathbf{A}}|} \log \left(p(\underline{\mathbf{x}}) \frac{1}{|\det \underline{\mathbf{A}}|} \right) |\det \underline{\mathbf{A}}|\, d\underline{\mathbf{x}} \\
&= -\int p(\underline{\mathbf{x}}) \log\, p(\underline{\mathbf{x}}) d\underline{\mathbf{x}} - (-\log\,|\det \underline{\mathbf{A}}|) \underbrace{\int p(\underline{\mathbf{x}}) d\underline{\mathbf{x}}}_{=1} \\
&= H(X) + \log\,|\det \underline{\mathbf{A}}|
\end{aligned}$$

$$\begin{aligned}
J(\underline{\mathbf{A}}\,\underline{X}) &= J(\underline{Y}) \\
&= H(\underline{Y}_G) - H(\underline{Y}) \\
&= \frac{1}{2}\log|\det\underline{\mathbf{\Sigma}}_Y| + \frac{N}{2}(1 + 2\log 2\pi) \quad - \quad \left[H(\underline{X}) + \log|\det\underline{\mathbf{A}}|\right] \\
&= \frac{1}{2}\log|\det\underline{\mathbf{A}}\,\underline{\mathbf{\Sigma}}\,\underline{\mathbf{A}}^\top| \quad + \quad \frac{N}{2}(1 + 2\log 2\pi) \quad - \quad H(\underline{X}) - \log|\det\underline{\mathbf{A}}| \\
&= \frac{1}{2}\log|\det\underline{\mathbf{\Sigma}}| + \frac{2}{2}\log|\det\underline{\mathbf{A}}| \quad + \quad \frac{N}{2}(1 + 2\log 2\pi) - H(\underline{X}) - \log|\det\underline{\mathbf{A}}| \\
&= \frac{1}{2}\log|\det\underline{\mathbf{\Sigma}}| + \frac{N}{2}(1 + 2\log 2\pi) - H(X) \\
&= H(\underline{X}_G) - H(\underline{Y}) \\
&= J(X)
\end{aligned}$$

i.e. negentropy is scale-invariant.