FastICA & the independent components of image patches

This problem sheet illustrates how to use approximations of *negentropy* (as implemented in the *FastICA* algorithm) to separate mixed signals. You can use an available FastICA implementation (fastICA is implemented in scikit-learn and in further toolboxes, cf. http://research.ics.aalto.fi/ica/fastica/).

7.1 Negentropy is scale-invariant (4 points)

The differential entropy of a n-dimensional random vector $\mathbf{x}$ with probability density $p(\mathbf{x})$ is defined as

$$H(\mathbf{x}) = -\int_{\mathbb{R}^n} p(\mathbf{x}) \log\, p(\mathbf{x}) d\mathbf{x}.$$

The negentropy is defined as

$$J(\mathbf{x}) = H(\mathbf{x}_{Gauss}) - H(\mathbf{x})$$

where $\mathbf{x}_{Gauss}$ is a n-dimensional multivariate Gaussian random vector with the same covariance matrix as $\mathbf{x}$.

Show that the negentropy is invariant w.r.t. invertible linear transformations $\mathbf{y} = \mathbf{A}\mathbf{x}$, i.e.

$$J(\mathbf{A}\mathbf{x}) = J(\mathbf{x})$$

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

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

$$H(\mathbf{x}_{Gauss}) = \frac{1}{2}\, \log\, |\det\, \mathbf{\Sigma}| + \frac{n}{2}\,(1 + \log\, 2\pi).$$

Remark: the differential entropy itself is not scale-invariant.

7.2 fastICA vs. Infomax (2 points)

Apply fastICA to the two soundfiles data set (once again) of problem sheet 6 and compare runtime and robustness w.r.t. the mixing matrix $\mathbf{A}$ with the Infomax-based ICA-algorithm. Use the following setup for the latter algorithm: natural gradient, Bell-Sejnowski amplitude normalization, learning rate schedule $\varepsilon_0 = 0.01$, $\varepsilon_{t+1} = 0.9999\varepsilon_t$.

7.3 ICA on Image Patches (4 points)

The file `imgpca.zip` (used also in exercise sheet 2) contains three categories of images: *nature*, *buildings*, and *text* (prefixes `n,b,t`). For each category:

(a) Sample P patches of $\sqrt{N} \times \sqrt{N}$ pixels from all images of this category and rearrange each sample to a column vector. Choose number and size of the patches according to your computing resources. Recommended are $P \geq 20000$ and $N \geq 144$.

(b) Calculate the independent features of the image patches (these are the columns of mixing matrix $\mathbf{A}$). Use a `fastICA` toolbox to compute this matrix:

- Let `fastica` perform PCA and whitening of the data.
- Use the contrast function $G(\hat{s}) = \frac{1}{a} \log\cosh(a\hat{s})$ with $a = 1$.

(c) Show the first 20 independent features as (grayscale) image patches by rearranging the vectors into $\sqrt{N} \times \sqrt{N}$ matrices and compare the results for the different categories. Order the independent features by decreasing negentropy, (such that the first feature has largest (approximated) negentropy etc).

(d) Perform PCA on the same set of patches, plot the the principal components (ordered by decreasing eigenvalue) as in (c) and compare them with the independent features.

Total points: 10