02458 Modelling Cognition / ICA exercise

Last time we saw how we could analyze the natural statistics of images using Principal Component Analysis to obtain a more efficient internal representation. This was entirely based on Gaussian statistics and you may have wondered if that was a reasonable assumption. In fact, when we look at natural statistics, we see that the distributions are characterized by higher moments than just the variance, which means that the Gaussian assumption is not a good one.

1. Let's have a look at what happens when do PCA on a bivariate distribution with excess kurtosis. First, pull 2 times 10000 samples from a Laplacian distribution using the randlpl.m function that comes with the exercise. Mix the two distributions, i.e. create two linear combinations of them. Plot the linear combinations against each other in a 2D plot.

2. Now perform PCA on the joint distribution. Plot the score in a 2D plot. Did PCA separate the two distributions?

3. Now get the fastICA toolbox (http://research.ics.tkk.fi/ica/fastica/) and perform ICA (score=fastica(data'))on the data. Be careful to flip your matrix the right way. Plot the ICA score in a 2D plot to see that the distribution now appear independent.

4. Compare the ICA score with the original samples. Are they identical? Why not? What can go wrong in recovering the original data?

5. Take two mono sound files with an equal amount of samples. You can use the guitar sound files in the file-sharing folder or you can use your own. Create two mixes of the sound files. Separate them using ICA.

    a. Plot (part of) the time series of the original sounds and of the reconstructed sounds. Estimate the error. Evaluate how well it worked.
    b. Listen to the sounds. Evaluate how well it worked based on your listening experience.
    c. Compare the two ways of evaluating whether ICA managed to separate the sounds. Why might the two methods arrive at different results?