# Foundations of Graphical Models: Homework 2

Out: T 2015-10-20
Due: W 2015-11-04

The total page limit for problems 1 and 2 is three pages (though you may use extra pages for figures.) Please use the latex template from the reading responses.

**Problem 1**

Implement Gibbs sampling for a Gaussian mixture (collapsed or uncollapsed) as we discussed in class. If you know how to do it, you can implement a different mixture model instead.

Apply your code to some data, either real or simulated, and discuss what you learned. You can plot and discuss whatever you like. Among the plots, we would like to see $\log p(x_{1:n}, z_{1:n}, \mu_{1:K})$ as a function of iteration. (This is a way to see how the Gibbs sampler converges.)

Getting through this exercise is important for having a good final project and, more generally, becoming fluent in the material. There are many "gotchas" in developing and deploying probabilistic models, which are only learned from experience. (For example, you will want to work in log space, only exponentiating when you need to. Maja will go over the "log-sum-exp" trick.)

**Problem 2**

This problem is intended to help you brainstorm ideas for the project. Consider some data. This can be the data you used in problem 5 of homework 1, or it can be other data. If you have a data set in mind for your final project then we encourage you to use it for this exercise as well.

a) **Variables in the data.** What are the variables in the data and what are some of their relationships to each other? Do you expect some of the variables to be correlated? Do you expect others to be (conditionally) independent?

b) **Latent variables.** What are some latent variables you could introduce to capture the correlations between model variables? What are some latent variables that could summarize aspects of the data? What other latent variables could be hidden in the data?

c) **Research question.** Formulate several questions you might be able to answer with the data. Write down the three most interesting ones.