Machine Learning 2 **Due: May 2, 2017**

Homework 4

Instructor: Stephan Bongers, Tineke Blom Email: s.r.bongers@uva.nl, t.blom2@uva.nl

You are allowed to discuss with your colleagues but you should write the answers in *your own words*. If you discuss with others, write down the name of your collaborators on top of the first page. No points will be deducted for collaborations. If we find similarities in solutions beyond the listed collaborations we will consider it as cheating.
We will not accept any late submissions under any circumstances. The solutions to the previous homework will be handed out in the class at the beginning of the next homework session. After this point, late submissions will be automatically graded zero.

Problem 1. Consider a Gaussian mixture model

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

1. Given the expected value of the complete-data log-likelihood (9.40 in Bishop's book)

$$\mathbb{E}_{\text{posterior}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left\{ \ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$$

   Derive update rules for $\boldsymbol{\pi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

2. Consider a special case of the model above, in which the covariance matrices $\boldsymbol{\Sigma}_k$ of the components are all constrained to have a common value $\boldsymbol{\Sigma}$. Derive EM equations for maximizing the likelihood function under such a model.

Problem 2. Suppose we wish to use the EM algorithm to maximize the posterior distribution

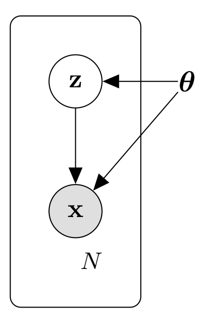

Figure 1: A simple generative model.

$p(\boldsymbol{\theta}|\mathbf{X})$ for a model (Figure 1) containing latent variables $\mathbf{z}$ and observed variables $\mathbf{x}$. Show that the E step remains the same as in the maximum likelihood case, where as in the M step, the quantity to be maximized is

$$\sum_{\mathbf{z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta})$$

Problem 3.

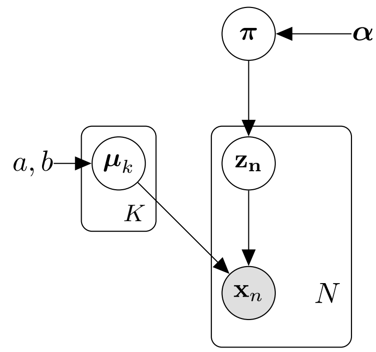

Figure 2: Mixtures of Bernoulli distribution

$$\begin{aligned}
\boldsymbol{\pi}|\boldsymbol{\alpha} &\sim \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}) \\
\mathbf{z}_n|\boldsymbol{\pi} &\sim \mathrm{Mult}(\mathbf{z}_n|\boldsymbol{\pi}) \\
\boldsymbol{\mu}_k|a_k, b_k &\sim \mathrm{Beta}(\boldsymbol{\mu}_k|a_k, b_k) \\
\mathbf{x}_n|\mathbf{z}_n, \boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K\} &\sim \prod_{k=1}^{K} (\mathrm{Bern}(\mathbf{x}_n|\boldsymbol{\mu}_k))^{z_{nk}}
\end{aligned}$$

Derive the EM algorithm for maximizing the posterior probability $p(\boldsymbol{\mu}, \boldsymbol{\pi}|\{\mathbf{x}_n\}_{n=1}^N)$. (The E step is given in Bishop's Book, you only need to do the M step)