Machine Learning 2 **Due: 23:59, October 7, 2019**

Homework 5

Instructors: Adeel Pervez, Pim de Haan, Noud de Kroon, David Zhang
Email: a.a.pervez@uva.nl, pim.de.haan@uva.nl, a.a.w.m.dekroon@uva.nl, w.d.zhang@uva.nl

Problem 1. (3 + 1 = 4 points)
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. (2 points)
Suppose we wish to use the EM algorithm to maximize the posterior distribution $p(\boldsymbol{\theta}|\mathbf{X})$ for a

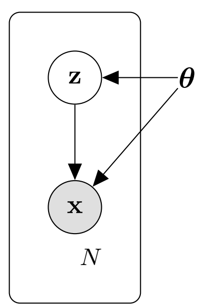

Figure 1: A simple generative model.

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. (3 points)

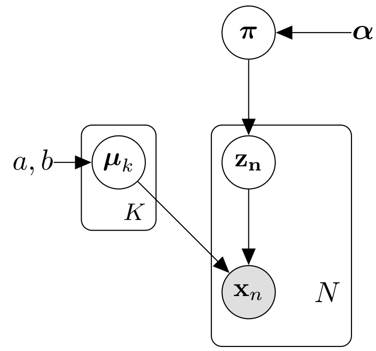

Figure 2: Mixtures of Bernoulli distribution

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

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)