STA561/CS571 — Fall 2013 **Midterm Exam** Due: October 16, 2013

Instructions: Please put all answers in a single PDF with your name and NetID and upload to SAKAI before class on the due date (there is a LaTeX template on the course web site for you to use). Do not work in a group; this is an individual exam. If you look at any references (even wikipedia), cite them in your answers.

Multivariate normal distribution, factor analysis, and EM

Please read the Ghahramani & Hinton paper on Sakai, "The EM algorithm for mixtures of factor analyzers", 1996. Note that the basic model, in Equation (1), is identical to our formulation of factor analysis for $x = x_i$, $z = z_i$ (a single sample $i \in \{1, \ldots, n\}$), and $u = \epsilon$. Throughout, you should include a sample index on the z variables as well as the x variables (Equations 1-4 should have i indices on both x and z, and in Equations 5 and 6, for example, put the same index i on the z variables).

1. To Equation (1), let's explicitly add a mean term across the p-dimensions, so that the equation (adding the sample index) becomes $x_i = \Lambda z_i + \mu + u$, where $\mu \in \Re^p$. As in the paper, we will assume $z_i \sim \mathcal{N}_k(0, I)$, where I is the identity matrix, and $u \sim \mathcal{N}_p(0, \Psi)$, where Ψ is a diagonal matrix with $(\psi_1, \cdots, \psi_p)$ variance terms on the diagonal. We are going to follow the steps in this paper to derive the EM algorithm for this adapted model.

   Throughout your response, write out all work. Do not skip steps, and do not make assumptions about what we know. If you are using knowledge about the problem other than simple algebra (e.g., an identity of the MVN distribution), please state this explicitly. The solution to this first problem should, if done correctly and completely, be multiple pages long.

   (a) Write out the expected log likelihood for this model (this will be a variant of the first equations in Appendix A). Determine the Expected sufficient statistics.

   (b) Derive the maximum likelihood estimates of μ, Ψ, and Λ using this equation, assuming that you have in hand values for the expected sufficient statistics.

   (c) Use the ideas behind Equations (3) and (4) to derive the expected sufficient statistics, assuming that you have in hand current values for the parameters.

   (d) Put this all together: Carefully write out the EM algorithm pseudocode, starting with how you might initialize your parameters, how you will set K, the number of factors, referring to the equations you derived in the earlier two steps.

   (e) How would you assess convergence of this EM algorithm?

2. In your EM algorithm, look carefully at the parameter estimates for the mean parameter μ. In two sentences, explain why modeling the mean values of each of the features in the model is different than mean-centering each of the features across the samples before performing factor analysis on this model?

3. Carefully describe a specific problem in which a mixture of factor analyzers would be preferable to a simple factor analysis model. Explain what the benefit of the mixture would be: describe the specific parameters, their interpretation in the problem you describe, and how their estimated values would allow you to solve the problem in a way that the factor analysis model would not.