CS 228T QUIZ 5

1. Give true/false answers, with brief justifications, to the following.

(a) In the fully observed setting, it is useful to do random restarts, *i.e.*, to run the parameter estimation algorithm multiple times from different starting points to obtain different estimated parameters, and then take the best one.
(b) In the partially observed setting, it is useful to do random restarts.

2. Suppose we are given an unlabeled training set $\mathcal{D} = \{x_1, \ldots, x_N\}$ and we wish to model the data using a mixture of K Gaussians:

$$\begin{aligned} z_n &\sim \text{Multinomial}(\theta) \\ x_n | z_n &\sim \text{Normal}(\mu_{z_n}, \Sigma_{z_n}). \end{aligned}$$

Here, the z_n are hidden variables that take on values in $\{1, \ldots, K\}$.

(a) State the optimization problem we want to solve to fit this model to the data. You don't need to bother expanding out the mass/density functions of the multinomial and Gaussian distributions.
(b) Concisely explain how to use EM to fit this model. You do not have to derive anything; just explain the outline of the method.
(c) Explain what hard EM looks like in this setting.

3. Here we consider the variational formulation of EM.

(a) In terms of the free energy functional, state the optimization problem that EM is (implicitly) solving via coordinate ascent.
(b) What is the main computational cost in solving this problem with EM?
(c) Briefly explain how a variational inference algorithm could be used to alleviate this problem. Does it matter whether we use loopy BP or mean field in this context?

4. This question considers Bayesian statistics and variational Bayes.

(a) What is the major difference between the way parameters are treated in Bayesian statistics and in classical statistical methods like maximum likelihood estimation?
(b) In the context of Bayesian statistics, why do *inference* algorithms (variational or otherwise) provide a full methodology for addressing the *learning* problem?
(c) What is the particular form of variational inference used in variational Bayes? In other words, which class of variational method is used and what is the exact form of the approximation? Explain why we might prefer this class of method to something else.

Date: September 29, 2020.