

*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). Definitely consider working in a group; please include the names of the people in your group and write up your solutions separately. If you look at any references (even wikipedia), cite them. If you happen to track the number of hours you spent on the homework, it would be great if you could put that at the top of your homework to give us an indication of how difficult it was.

### Problem 1

#### *Expectation Propagation*

- (a) Consider a factorized approximation of  $q(x, y) = q(x)q(y)$  for joint distribution  $p(x, y)$ . Minimize, with respect to  $q(x)$  (and, identically, for  $q(y)$ , no need to write both out), the *forward* KL divergence,  $KL(p \parallel q)$ . (Hint: use the standard approach of taking the derivative of our new variational objective with respect to  $q(x)$  and setting to zero. When you get stuck, remember that you should include the constraint that  $\sum_x q(x) = 1$  in the form of a Lagrange multiplier).
- (b) Write two sentences that capture the variational updates implied from this approach.

### Problem 2

#### *Dirichlet distribution*

In a number of the models we are about to look at in class, including the latent Dirichlet allocation (LDA) model and the Dirichlet process mixture model, the parameterization of a Dirichlet,  $\alpha$ , plays a substantial role in the clustering effect of the model. Recall that the Dirichlet distribution is the conjugate prior for the parameters of a multinomial. In particular, let

$$\begin{aligned}\pi_j &= \text{Dir}(\alpha) \\ z_{i,j} &= \text{Mult}(\pi_j) \\ x_{i,j} &= \mathcal{N}(\mu_k, \sigma_k^2).\end{aligned}$$

For each object  $j = 1, \dots, p$ , we generate object-specific mixture proportions  $\pi_j$ , and, from this, we generate an element-specific class  $z_{i,j}$  and univariate normal value  $x_{i,j}$ , where  $i = 1, \dots, n$ .

- (a) When  $\alpha = [\alpha_1, \dots, \alpha_K]$  are all 0.2, what does this imply about the resulting clustering of elements within each object  $j$ ?
- (b) When  $\alpha = 1$ , what does this imply about the clustering of elements within each object  $j$ ?

- (c) When  $\alpha = 10$ , what does this imply about the clustering of elements within each object  $j$ ?
- (d) For  $K = 2$ , a Dirichlet distribution is captured in a *beta* distribution. Plot a histogram of samples from a  $\text{beta}(0.2, 0.2)$ ,  $\text{beta}(1, 1)$ , and  $\text{beta}(10, 10)$ . How do these support your answers above?
- (e) Looking at these histograms, do you have an intuition for the settings of  $\alpha$  for which inference in these models is easier? What is that setting, and why might you suspect that inference is easier?

**Problem 3***Project progress*

Write a few sentences on your work this week on your course project. What roadblocks have you hit? Are there any questions about specific approaches? Did you get an exciting result you can share with us? Tell us about it.