CS 228, Winter 2012
Theory Problem #5

Due: 12:00 noon on Friday, March 16th. **NO LATE DAYS**

EM for Data Association

Consider a data association problem where we have a set of objects $\mathcal{U} = \{u_1, \ldots, u_K\}$, of which we obtain several rounds of observations. That is, at each round of observation $r = 1, \ldots, R$, we obtain a set of measurements $v_1^r, \ldots, v_K^r$, where each observation corresponds to one object and there is only one observation of each object. We capture this by introducing a set of correspondence variables $\mathcal{C} = \{C_1^1, \ldots, C_K^1, C_1^2, \ldots C_K^R\}$ such that $Val(C_i^r) \in \{1, \ldots, K\}$ for all i, r. We have that $C_i^r = k$ if measurement v_i^r is derived from object u_k. The observation rounds occur at non-consecutive time points, so that there are no correlations between the correspondences at different time points. Thus, *a priori*, each measurement is equally likely to arise from each object. However, each object u has a certain appearance (e.g., color), and a measurement that is derived from u is likely to match that. Specifically, we assume that each measurement v_i^r is some discrete value in some space $A = \{a_1, \ldots, a_L\}$ (e.g., the space of colors). We model the appearance of u as a multinomial distribution $\theta^u = (\theta_1^u, \ldots, \theta_L^u)$ over the set A. If measurement v_i^r comes from object u_k, that is, if $C_i^r = k$, then $v_i^r = a_\ell$ with probability $\theta_\ell^{u_k}$. In this question we will ask you to describe how you would use EM to learn the appearance model θ^u for all objects u.

(a) **[10 points]** A good first step when figuring out how to apply EM to a problem is to determine the sufficient statistics. Write out the likelihood function and sufficient statistics in the case of fully observed data (objects, correspondences, and measurements are observed) for this problem. What are the sufficient statistics for learning θ^u in the general data association problem where $\mathcal{C}$ is not observed? Please introduce notation that you can use consistently throughout the question.

(b) **[10 points]** Specify how you would do maximum likelihood estimation of the parameters in the complete data case. Given the sufficient statistics described in the previous problem, describe how the parameters would be estimated in the M-step for the general data association problem described above (using MLE).

(c) **[10 points]** Finally, define the inference task that you would need to solve in the E-step: in what probabilistic model you would run inference in order to derive the sufficient statistics. (There is no need to discuss specific inference algorithms.) Is the E-step in this case computationally easy or hard?

(d) **[10 points]** Now we want to solve the same problem using a hard EM variant, in which, in each iteration, rather than doing a soft assignment of the hidden variables to their possible values, we use the MAP assignment in each instance, and then treat the hidden variables as observed to take these values (for this iteration). Describe how hard EM would be used in the context of this algorithm. Is this computationally harder or easier than soft EM in this case?