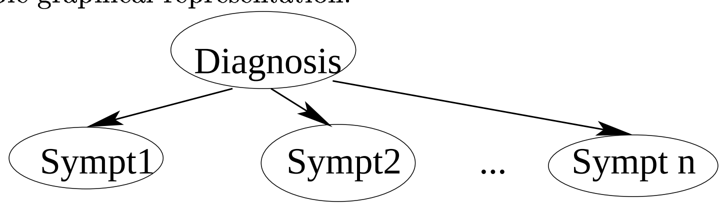
Probabilistic Reasoning in AI - Assignment 5

Due Friday, March 26, in class

1. [60 points] **Expectation maximization (EM)**

   Once upon a time, at the beginning of the term, we discussed the naive Bayes model, which has the following simple graphical representation:

   In this problem, you will derive the EM algorithm for learning this model from data, assuming that the diagnosis is not always observed.

   Let $S_1, \dots S_n$ be the random variables denoting the symptoms and D be the random variable denoting the diagnosis. We will assume that D can take K possible values and S_i can each take L possible values, $s_{i1}, \dots s_{iL}$. The parameters of the model are:

   $$a_k = p(D = d_k), k = 1, \dots K b_{ilk} = p(S_i = s_{il} | D = d_k), i = 1, \dots n, l = 1, \dots L, k = 1, \dots K$$

   (a) [5 points] Suppose you have a complete data set of m instances, of the form: $\langle s_1^j, \dots s_n^j, d^j \rangle, j = 1, \dots m$. Write down the likelihood of the parameters given the data.

   (b) [5 points] Compute the maximum likelihood parameters for the model, given complete data.

   (c) [10 points] Now assume that in p out of the m instances, the value of the D random variable is unobserved. Calculate the marginal log likelihood of the data in this case.

   (d) [10 points] E-step: Suppose the jth instance is missing the diagnosis. Compute $c_k^j = p(d_k^j | s_1^j, \dots s_n^j), k = 1, \dots K$.

   (e) [5 points] Compute the expected complete log likelihood of the data set, using the c^js computed above.

   (f) [15 points] M-step: given $c_k^j, k = 1, \dots K$, and given the data, compute the *best* setting for the parameters of the model, a_k^* and b_{ilk}^*. Make sure you enforce the condition that probabilities sum to 1: $\sum_k a_k^* = 1$ and $\sum_l b_{ilk}^* = 1 \forall i, k$.

   (g) [10 points] Based on your answer to the previous questions, write down an EM algorithm for computing the model parameters from the data. The E-step will updated the cs, and the M-step will updated the as and bs.

2. [15 points] **Hidden Markov Models**

   In class we discussed the forward-backward algorithm in the context of inferring $p(y_1, \dots y_t, s_t)$, where s_t is the state at time t and $y_1, \dots y_t$ are observations. Suppose that instead, we define $\alpha = p(s_t | y_1, \dots y_t)$. Derive the update rule for α in this case. Express the update rule in terms of matrix-vector multiplications.

3. [15 points] **Problem formulation** Consider a simplified robot localization problem, in which Roby the robot is going around the 3rd floor in McConnell, trying to figure out where it is. Roby has a map of the floor, and a model of how much it can travel in a second. It also has a sensor which, at every time step, tells it whether it is in a hallway or at an intersection. Unfortunately, the sensor malfunctions 10% of the time, in which case it gives Roby the wrong reading. Roby always starts in front of room 318, and its trajectories are generated based on a fixed controller.

   Describe in detail a probabilistic model that can be used to model this scenario. Describe what inference method you would use to help Roby localize.

4. [10 points] **Structural EM**

   We discussed in class the structural EM algorithm for inferring the structure of Bayes nets. Suppose you have a Bayes net with some variables $X_1, \dots X_n$ which are always observed, and a variable H which is hidden *all the time*. Suppose you start with a structure in which H is not connected to any of the observed variables. The variables $X_1, \dots X_n$ are connected in some arbitrary way. The structural EM algorithm will complete the data with values for H. It will then consider adding or removing arcs, and it will evaluate these "successor" candidates based on the complete data. Let B' be the best candidate according to the score. Show that in B', H must be disconnected from $X_1, \dots X_n$.