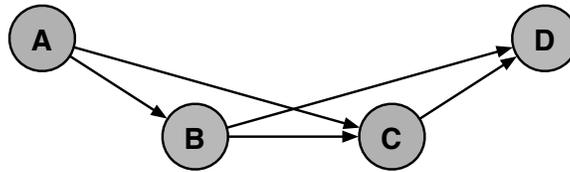


**Out:** Tue Aug 19

**Due:** Tue Aug 26

### 5.1 Maximum likelihood estimation



(a) **Complete data**

Consider a complete data set of *i.i.d.* examples  $\{a_t, b_t, c_t, d_t\}_{t=1}^T$  drawn from the joint distribution of the above belief network. Compute the maximum likelihood estimates of the conditional probability tables (CPTs) shown below for this data set. Express your answers in terms of indicator functions, such as:

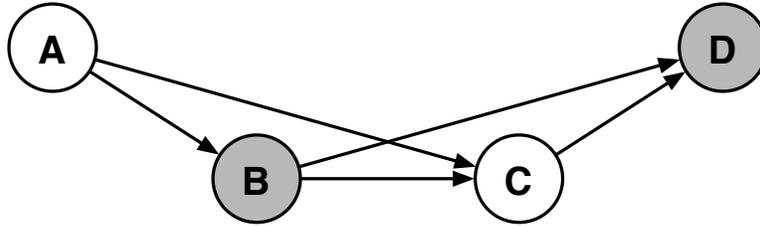
$$I(a, a_t) = \begin{cases} 1 & \text{if } a = a_t, \\ 0 & \text{if } a \neq a_t. \end{cases}$$

For example, in terms of this indicator function, the maximum likelihood estimate for the CPT at node  $A$  is given by  $P(A=a) = \frac{1}{T} \sum_{t=1}^T I(a, a_t)$ . Complete the numerators and denominators in the below expressions.

$$P(B=b|A=a) = \underline{\hspace{4cm}}$$

$$P(C=c|A=a, B=b) = \underline{\hspace{4cm}}$$

$$P(D=d|B=b, C=c) = \underline{\hspace{4cm}}$$



(b) **Posterior probability**

Consider the belief network shown above, with observed nodes  $B$  and  $D$  and hidden nodes  $A$  and  $C$ . Compute the posterior probability  $P(a, c|b, d)$  in terms of the CPTs of the belief network—that is, in terms of  $P(a)$ ,  $P(b|a)$ ,  $P(c|a, b)$  and  $P(d|b, c)$ .

(c) **Posterior probability**

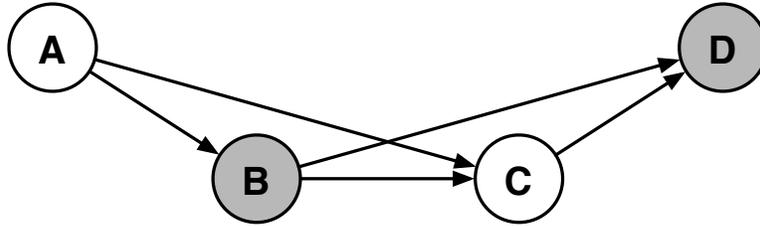
Compute the posterior probabilities  $P(a|b, d)$  and  $P(c|b, d)$  in terms of your answer from part (b). In other words, in this problem, you may assume that  $P(a, c|b, d)$  is given.

(d) **Log-likelihood**

Consider a partially complete data set of *i.i.d.* examples  $\{b_t, d_t\}_{t=1}^T$  drawn from the joint distribution of the above belief network. The log-likelihood of the data set is given by:

$$\mathcal{L} = \sum_t \log P(B = b_t, D = d_t).$$

Compute this log-likelihood in terms of the CPTs of the belief network. You may re-use work from earlier parts of the problem.



(e) **EM algorithm**

The posterior probabilities from part (b) can be used by an EM algorithm to estimate CPTs that maximize the log-likelihood from part (c). Complete the numerator and denominator in the below expressions for the EM update rules. Simplify your answers as much as possible, expressing them in terms of the posterior probabilities  $P(a, c|b_t, d_t)$ ,  $P(a|b_t, d_t)$ , and  $P(c|b_t, d_t)$ , as well as the indicator functions  $I(b, b_t)$ , and  $I(d, d_t)$ .

$P(A=a) \leftarrow$  \_\_\_\_\_

$P(B=b|A=a) \leftarrow$  \_\_\_\_\_

$P(C=c|A=a, B=b) \leftarrow$  \_\_\_\_\_

$P(D=d|B=b, C=c) \leftarrow$  \_\_\_\_\_

---

## 5.2 EM algorithm for noisy-OR

Consider the belief network on the right, with binary random variables  $X \in \{0, 1\}^n$  and  $Y \in \{0, 1\}$  and a noisy-OR conditional probability table (CPT). The noisy-OR CPT is given by:

$$P(Y = 1|X) = 1 - \prod_{i=1}^n (1 - p_i)^{X_i},$$

which is expressed in terms of the noisy-OR parameters  $p_i \in [0, 1]$ .

In this problem, you will use the EM algorithm derived in class for estimating the noisy-OR parameters  $p_i$ . For a data set  $\{(\vec{x}_t, y_t)\}_{t=1}^T$ , the (normalized) conditional log-likelihood is given by:

$$\mathcal{L} = \frac{1}{T} \sum_{t=1}^T \log P(Y = y_t | X = \vec{x}_t).$$

Download the data files on the course web site, and use the EM algorithm to estimate the parameters  $p_i$ . The data set has  $T=267$  examples over  $n=23$  inputs. For those interested, more information about this data set is available here:

<http://archive.ics.uci.edu/ml/datasets/SPECT+Heart>

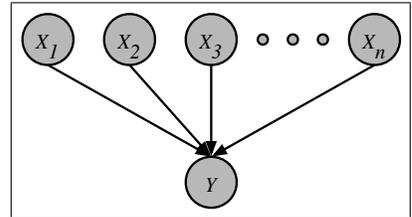
**However, be sure to use the data files provided on the course web site, as they have been specially assembled for this assignment.** The EM update for this model is given by:

$$p_i \leftarrow \frac{1}{T_i} \sum_{t=1}^T \frac{y_t x_{it} p_i}{1 - \prod_{j=1}^n (1 - p_j)^{x_{jt}}},$$

where  $T_i$  is the number of examples in which  $X_i = 1$ . Initialize all  $p_i = \frac{1}{n}$  and perform 512 iterations of the EM algorithm. At each iteration, compute the conditional log-likelihood shown above. (If you have implemented the EM algorithm correctly, this conditional log-likelihood will always increase from one iteration to the next.) Also compute the number of mistakes  $M \leq T$  made by the model at each iteration; a mistake occurs either when  $y_t = 0$  and  $P(y_t = 1 | \vec{x}_t) \geq 0.5$  (indicating a false positive) or when  $y_t = 1$  and  $P(y_t = 1 | \vec{x}_t) \leq 0.5$  (indicating a false negative). The number of mistakes should generally decrease as the model is trained, though it is not guaranteed to do so at each iteration.

**Turn in all of the following:**

- a hard-copy print-out of your source code
- a plot or print-out of your final estimated values for  $p_i$
- a completed version of the following table



| iteration | number of mistakes $M$ | conditional log-likelihood $\mathcal{L}$ |
|-----------|------------------------|--|
| 0         | 195                    | -1.045                                   |
| 1         | 60                     |  |
| 2         |                        | -0.4108                                  |
| 4         |                        |  |
| 8         |                        |  |
| 16        |                        |  |
| 32        |                        |  |
| 64        |                        |  |
| 128       | 36                     |  |
| 256       |                        |  |
| 512       |                        | -0.3100                                  |

You should use the already completed entries of this table to check your work. As always you may program in the language of your choice.

---