Optimization and Machine Learning, Spring 2021
Homework 3

(Due Thursday, Apr. 22 at 11:59pm (CST))

RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 . . . 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 . . . 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 . . . 40	medium	no	excellent	yes
13	31 . . . 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Figure 1: The Buy Computer data.

1. [15 points] Given the training data in Fig. 1, please use Naive Bayes classifier to predict the class of the following new example: age<= 30, income=medium, student=yes, credit-rating=fair.
Solution:

E= age<=30, income=medium, student=yes, credit-rating=fair
E_1 is age<=30, E2 is income=medium, student=yes, E4 is credit-rating=fair
We need to compute P(yes|E) and P(no|E) and compare them.

$$P(yes \mid E) = \frac{P(E_1 \mid yes)\,P(E_2 \mid yes)\,P(E_3 \mid yes)\,P(E_4 \mid yes)\,P(yes)}{P(E)}$$

P(yes)=9/14=0.643 P(no)=5/14=0.357

P(E1|yes)=2/9=0.222 P(E1|no)=3/5=0.6
P(E2|yes)=4/9=0.444 P(E2|no)=2/5=0.4
P(E3|yes)=6/9=0.667 P(E3|no)=1/5=0.2
P(E4|yes)=6/9=0.667 P(E4|no)=2/5=0.4

$$P(yes \mid E) = \frac{0.222\ 0.444\ 0.667\ 0.668\ 0.443}{P(E)} = \frac{0.028}{P(E)} \qquad P(no \mid E) = \frac{0.6\ 0.4\ 0.2\ 0.4\ 0.357}{P(E)} = \frac{0.007}{P(E)}$$

Hence, the Naïve Bayes classifier predicts buys_computer=yes for the new example.

2. [20 points] Given the Bayesian network shown in Fig. 2, answer the following questions.

(a) Use the D-separation to analyze the two cases: [10 points]

- Given X_4, $\{X_1, X_2\}$ and $\{X_6, X_7\}$ are conditionally independent.
- Given $\{X_6, X_7\}$, X_3 and X_5 are conditionally independent.

Solution:

- The statement is True. According to D-separation, $\{x_1, x_2\}$ and $\{x_6, x_7\}$ can be regarded as two sets A and B. All the arrows on the path form A to B meet head-to-tail, therefore all the paths are blocked given x_4.
- The statement is False. The arrow on the path from x_3 to x_4 meets head-to-head. Since the node x_6 is observed, the path from x_3 to x_4 is not blocked. The path from x_3 to x_4 is the same. The path from x_6 to x_7 is also unblocked, therefore x_3 and x_5 are not conditionally independent.

(b) If all the nodes in Fig. 2 are observed and boolean variables, please learn the parameter $\theta_{4|i,j}$ according to Maximum Likelihood Estimation (MLE), where $\theta_{4|i,j} = p(X_4 = 1 \mid X_1 = i, X_2 = j), i, j \in \{0, 1\}$. [10 points]

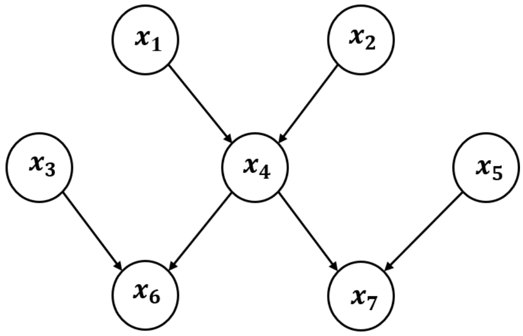

Figure 2: A Bayesian network with seven nodes.

Solution: Suppose we observed K data points. Let $\theta = \{\theta_{x_1}, \theta_{x_2}, \theta_{x_3}, \theta_{x_5}, \theta_{x_4|i,j}, \theta_{x_6|i,j}, \theta_{x_7|i,j}\}$, then

$$\begin{aligned}
\log p(\mathcal{D} \mid \theta) &= \log \prod_{k=1}^{K} p(x_{1k}, x_{2k}, x_{3k}, x_{4k}, x_{5k}, x_{6k}, x_{7k} \mid \theta) \\
&= \log \prod_{k=1}^{K} p(x_{1k} \mid \theta) p(x_{2k} \mid \theta) p(x_{3k} \mid \theta) p(x_{5k} \mid \theta) p(x_{4k} \mid x_{1k}, x_{2k}, \theta) p(x_{6k} \mid x_{3k}, x_{4k}, \theta) p(x_{7k} \mid x_{4k}, x_{5k}, \theta) \\
&= \sum_{k=1}^{K} \log p(x_{1k} \mid \theta) + \log p(x_{2k} \mid \theta) + \log p(x_{3k} \mid \theta) + \log p(x_{5k} \mid \theta) + \log p(x_{4k} \mid x_{1k}, x_{2k}, \theta) \\
&\quad + \log p(x_{6k} \mid x_{3k}, x_{4k}, \theta) + \log p(x_{7k} \mid x_{4k}, x_{5k}, \theta).
\end{aligned}$$

Then we derive the gradient of $\log p(\mathcal{D} \mid \theta)$ with respect to $\theta_{x_4|i,j}$

$$\frac{\partial \log p(\mathcal{D} \mid \theta)}{\partial \theta_{x_4|i,j}} = \sum_{ik=1}^{K} \frac{\partial p(x_{4k} \mid x_{1k}, x_{2k}, \theta)}{\partial \theta_{x_4|i,j}}$$

Set the derivative to 0 and then obtain the parameter $\theta_{x_4|i,j}$

$$\theta_{x_4|i,j} = \frac{\sum_{k=1}^{K} \mathbb{I}(x_{4k} = 1, x_{1k} = i, x_{2k} = j)}{\sum_{k=1}^{K} \mathbb{I}(x_{1k} = i, x_{2k} = j)},$$

where $\mathbb{I}(\cdot)$ is the indicator function.

3. [15 points] Given a Bayesian network (Fig. 3) with five discrete variables $\{F, A, S, H, N\}$, where $\{F, A, H, N\}$ are boolean variables and $S \in \{0, 1, 2\}$. Suppose that $\{F, A, N\}$ are observed variables and $\{S, H\}$ are latent variables. Now we implement EM algorithm for this model.

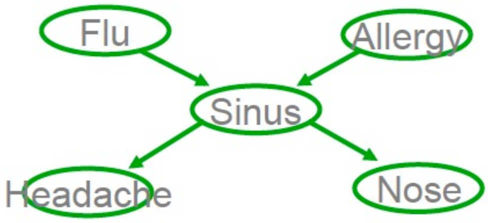

Figure 3: The Bayesian network with five discrete variables $\{F, A, S, H, N\}$.

(a) Derive the E-step. [5 points]
Solution: In E-step, calculate $P(S, H|F, A, N, \theta)$.

$$P(s_k = 0, h_k = 0|f_k, a_k, n_k, \theta) = \frac{P(s_k = 0, h_k = 0, f_k, a_k, n_k|\theta)}{\sum_{i=0}^{2}\sum_{j=0}^{1} P(s_k = i, h_k = j, f_k, a_k, n_k|\theta)},$$
$$P(s_k = 0, h_k = 1|f_k, a_k, n_k, \theta) = \frac{P(s_k = 0, h_k = 1, f_k, a_k, n_k|\theta)}{\sum_{i=0}^{2}\sum_{j=0}^{1} P(s_k = i, h_k = j, f_k, a_k, n_k|\theta)},$$
$$P(s_k = 1, h_k = 0|f_k, a_k, n_k, \theta) = \frac{P(s_k = 1, h_k = 0, f_k, a_k, n_k|\theta)}{\sum_{i=0}^{2}\sum_{j=0}^{1} P(s_k = i, h_k = j, f_k, a_k, n_k|\theta)},$$
$$P(s_k = 1, h_k = 1|f_k, a_k, n_k, \theta) = \frac{P(s_k = 1, h_k = 1, f_k, a_k, n_k|\theta)}{\sum_{i=0}^{2}\sum_{j=0}^{1} P(s_k = i, h_k = j, f_k, a_k, n_k|\theta)},$$
$$P(s_k = 2, h_k = 0|f_k, a_k, n_k, \theta) = \frac{P(s_k = 2, h_k = 0, f_k, a_k, n_k|\theta)}{\sum_{i=0}^{2}\sum_{j=0}^{1} P(s_k = i, h_k = j, f_k, a_k, n_k|\theta)},$$
$$P(s_k = 2, h_k = 1|f_k, a_k, n_k, \theta) = \frac{P(s_k = 2, h_k = 1, f_k, a_k, n_k|\theta)}{\sum_{i=0}^{2}\sum_{j=0}^{1} P(s_k = i, h_k = j, f_k, a_k, n_k|\theta)}.$$

(b) Derive the M-step. [5 points]
Solution: In M-step, choose θ' which maximize $E_{P(S,H|F,A,N,\theta)} \log P(S, H, F, A, N|\theta')$, where

$$E_{P(S,H|F,A,N,\theta)} \log P(S, H, F, A, N|\theta')$$
$$= \sum_{k=1}^{K}\sum_{i=0}^{2}\sum_{j=0}^{1} P(s_k = i, h_k = j|f_k, a_k, n_k, \theta)[\log P(f_k) + \log P(a_k) + \log P(s_k|f_k, a_k) + \log P(h_k|s_k) + \log P(n_k|s_k)].$$

(c) Guess the solution of parameter estimation in the M-step, according to the MLE solution with all variables being observed. [5 points]
Solution: The solutions of MLE are:

$$\theta_f = \frac{\sum_{k=1}^{K} \delta(f_k = 1)}{K},$$
$$\theta_a = \frac{\sum_{k=1}^{K} \delta(a_k = 1)}{K},$$
$$\theta_{s|f,a} = \frac{\sum_{k=1}^{K} \delta(s_k = s, f_k = f, a_k = a)}{\sum_{k=1}^{K} \delta(f_k = f, a_k = a)},$$
$$\theta_{h|s} = \frac{\sum_{k=1}^{K} \delta(h_k = 1, s_k = s)}{\sum_{k=1}^{K} \delta(s_k = s)},$$
$$\theta_{n|s} = \frac{\sum_{k=1}^{K} \delta(n_k = 1, s_k = s)}{\sum_{k=1}^{K} \delta(s_k = s)}.$$

By replacing $\delta(\cdot)$ by $P(\cdot)$ for the unobserved variables $\{S, H\}$, we have the solutions of the M-step in the EM algorithm:

$$\theta_f = \frac{\sum_{k=1}^{K} \delta(f_k = 1)}{K},$$
$$\theta_a = \frac{\sum_{k=1}^{K} \delta(a_k = 1)}{K},$$
$$\theta_{s|f,a} = \frac{\sum_{k=1}^{K} P(s_k = s)\delta(f_k = f, a_k = a)}{\sum_{k=1}^{K} \delta(f_k = f, a_k = a)},$$
$$\theta_{h|s} = \frac{\sum_{k=1}^{K} P(h_k = 1)P(s_k = s)}{\sum_{k=1}^{K} P(s_k = s)},$$
$$\theta_{n|s} = \frac{\sum_{k=1}^{K} \delta(n_k = 1)P(s_k = s)}{\sum_{k=1}^{K} P(s_k = s)}.$$