Istanbul Technical University- Spring 2017
BLG527E Machine Learning
Homework 4

Purpose: Graphical Models, Hidden Markov Models.

Total worth: 6% of your grade.

Handed out: Thursday, Dec 7, 2017.

Due: Thursday, Dec 22, 2017 23.00. (through ninova!)

Instructor: Zehra Cataltepe (cataltepe@itu.edu.tr),

Assistant: Mahiye Uluyağmur- Öztürk (muluyagmur@itu.edu.tr)

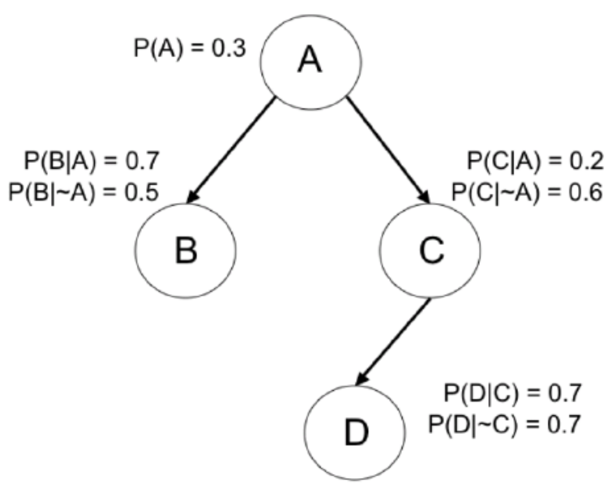
Policy: Collaboration in the form of discussions is acceptable, but you should write your own answer/code by yourself. Cheating is highly discouraged for it could mean a zero or negative grade from the homework. If a question is not clear, please let us know (via email, during office hour or in class).

Submission Instructions: Please submit through the class ninova site.
Write your answers to a report and upload it as a pdf file.

QUESTIONS:

Q1) [3 points] For the Bayesian network shown below, compute the following:

a) **[1 points]** P(A,B,C,D)=?
b) **[1 points]** P(A|B) =?
c) **[1 points]** P(C|B) =?



Q1_Solution)

a) P(A,B,C,D) =P(D|C)P(C|A)P(B|A)P(A) = 0.7 * 0.2 * 0.7 * 0.3 = 0.0294

b) P(A|B) = P(B,A)/P(B) =P(B|A)P(A)/(P(B|A)P(A)+P(B|~A)P(~A)) =0.7*0.3/(0.7*0.3+0.5*0.7) =0.21/0.56 =0.375

c) P(C|B) = $\sum_{A}$ P(C,A|B)=P(A|B).P(C|A)+P(~A|B).P(C|~A)=0,45

Q2) [3 points] You are given the following HMM with N=2 hidden states: S1, S2, M=2 possible observations: a,b, and state transition probabilities (A) and observation probabilities (B) and initial state probabilities (P).

a) **[1.5 points]** Compute the probability that the observation sequence O = a,a,b was produced by this HMM.

b) **[1.5 points]** What is the most probable state sequence given O?

$$A = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix} \quad B = \begin{array}{c} \begin{array}{cc} a & b \end{array} \\ \begin{bmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{bmatrix} \end{array} \qquad P = [0.9, 0.1]$$

Hint: The forward and backward variables in an HMM are calculated as it was follows:

Forward variable:

$$a_t(i) \equiv P(O_1 ... O_t, q_t = S_i | \lambda)$$

Initialization:

$$a_t(i) = \pi_i b_i(O_1)$$

Recursion:

$$a_{t+1}(j) = \left[\sum_{i=1}^{N} a_t(i) a_{ij} \right] b_j(O_{t+1})$$

$$P(O|\lambda) = \sum_{i=1}^{N} a_T(i)$$

Backward variable:

$$\beta_t(i) \equiv P(O_{t+1} ... O_T | q_t = S_i, \lambda)$$

Initialization:

$$\beta_T(i) = 1$$

Recursion:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$$

Q2_Solution)

a)

$\boldsymbol{P(O=\{a,a,b\} \mid \lambda) = ?}$

Let $\alpha_t(i)$ be probability of partial observation sequence <u>up to time</u> t where underlying Markov process is in state q_i at time t

$$\alpha_t(i) = P(O_0, O_1, \ldots O_{t-1}, x_t = q_i | \lambda) \quad \text{for } t=0,1, \ldots, T-1 \text{ and } i=0,1, \ldots, N-1$$

1. $\alpha_0(i) = \pi_i . b_i(O_0 = \mathrm{a})$ for $i=0,1$

$\alpha_0(0) = \pi_0 . b_0(O_0 = \mathrm{a}) = 0.9 x 0.1 = 0.09 \qquad \alpha_0(1) = \pi_1 . b_1(O_0 = \mathrm{a}) = 0.1 x 0.9 = 0.09$

2. $\alpha_t(i) = \left[\sum_{j=0}^{N-1} \alpha_{t-1}(j) . a_{ji}\right] . b_i(O_t)$ for $t=1,2$ and $i=0,1$

$$\alpha_1(0) = \left[\sum_{j=0}^{1} \alpha_0(j) . a_{j0}\right] . b_0(O_1 = a) = [\alpha_0(0) . a_{0,0} + \alpha_0(1) . a_{1,0}] . b_0(O_1 = a)$$
$$= [0.09 x 0.8 + 0.09 x 0.2] x 0.1 = \mathbf{0.009}$$

$$\alpha_1(1) = \left[\sum_{j=0}^{1} \alpha_0(j) . a_{j1}\right] . b_1(O_1 = a) = [\alpha_0(0) . a_{0,1} + \alpha_0(1) . a_{1,1}] . b_1(O_1 = a)$$
$$= [0.09 x 0.2 + 0.09 x 0.8] x 0.9 = \mathbf{0.081}$$

$$\alpha_2(0) = \left[\sum_{j=0}^{1} \alpha_1(j) . a_{j0}\right] . b_0(O_2 = b) = [\alpha_1(0) . a_{0,0} + \alpha_1(1) . a_{1,0}] . b_0(O_2 = b)$$
$$= [0.009 x 0.8 + 0.081 x 0.2] x 0.9 = \mathbf{0.02106}$$

$$\alpha_2(1) = \left[\sum_{j=0}^{1} \alpha_1(j) . a_{j1}\right] . b_1(O_2 = b) = [\alpha_1(0) . a_{0,1} + \alpha_1(1) . a_{1,1}] . b_1(O_2 = b)$$
$$= [0.009 x 0.2 + 0.081 x 0.8] x 0.1 = \mathbf{0.00666}$$

3. $P(O|\lambda) = \sum_{i=0}^{N-1} \alpha_{T-1}(i) = \sum_{i=0}^{1} \alpha_2(i) = \alpha_2(0) + \alpha_2(1) = 0.02106 + 0.00666 = \mathbf{0.02772}$

b)

Method - 1: α, β pass

Let $\beta_i(t)$ be probability of partial observation sequence <u>after the time</u> t where underlying Markov process is in state q_i at time t

$$\beta_i(t) = P(O_{t+1}, O_{t+2}, \dots O_{T-1}, x_t = q_i | \lambda) \quad \text{for } t = 0, 1, \dots, T-1 \text{ and } i = 0, 1, \dots, N-1$$

1. $\beta_{T-1}(i) = 1, \text{ for } i = 0, 1$

$\beta_2(0) = 1 \text{ and } \beta_2(1) = 1$

2. $\beta_t(i) = \sum_{j=0}^{N-1} a_{ij}.b_j(O_{t+1}).\beta_{t+1}(j) \quad \text{for } t = 1, 0 \text{ and } i = 0, 1$

$$\beta_1(0) = \sum_{j=0}^{1} a_{0j}.b_j(O_2=\text{b}).\beta_2(j) = a_{00}.b_0(O_2=\text{b}).\beta_2(0) + a_{01}.b_1(O_2=\text{b}).\beta_2(1)$$
$$= 0.8 x 0.9 x 1 + 0.2 x 0.1 x 1 = \mathbf{0.74}$$

$$\beta_1(1) = \sum_{j=0}^{1} a_{1j}.b_j(O_2=\text{b}).\beta_2(j) = a_{10}.b_0(O_2=\text{b}).\beta_2(0) + a_{11}.b_1(O_2=\text{b}).\beta_2(1)$$
$$= 0.2 x 0.9 x 1 + 0.8 x 0.1 x 1 = \mathbf{0.26}$$

$$\beta_0(0) = \sum_{j=0}^{1} a_{0j}.b_j(O_1=\text{a}).\beta_1(j) = a_{00}.b_0(O_1=\text{a}).\beta_1(0) + a_{01}.b_1(O_1=\text{a}).\beta_1(1)$$
$$= 0.8 x 0.1 x 0.74 + 0.2 x 0.9 x 0.26 = \mathbf{0.1060}$$

$$\beta_0(1) = \sum_{j=0}^{1} a_{1j}.b_j(O_1=\text{a}).\beta_1(j) = a_{10}.b_0(O_1=\text{a}).\beta_1(0) + a_{11}.b_1(O_1=\text{a}).\beta_1(1)$$
$$= 0.2 x 0.1 x 0.74 + 0.8 x 0.9 x 0.26 = \mathbf{0.2020}$$

Let $\gamma_t(i)$ be probability of Markov process being in state q_i at time t given observation sequence O. Since $\alpha_i(t)$ measures relevant probability up to time t and $\beta_i(t)$ measures the relevant probability after time t

$$\gamma_t(i) = P(x_t = q_i \mid O, \lambda) = \frac{\alpha_t(i).\beta_t(i)}{P(O \mid \lambda)}$$

Because denominator is common for all probabilities, it can be eliminated. So;

$$\gamma_t(i) = \alpha_t(i).\beta_t(i)$$

The most likely state at time t is the state q_i for which $\gamma_t(i)$ is maximum for index i

$$
\begin{array}{lllllllll}
\gamma_0(0) & = & \alpha_0(0).\beta_0(0) & = & 0.09 x 0.1060 & = & 0.00954 & & \\
\gamma_0(1) & = & \alpha_0(1).\beta_0(1) & = & 0.09 x 0.2020 & = & \mathbf{0.01818} & & \text{*Most probable state is } S_1 \text{ at time } t_0 \\
\gamma_1(0) & = & \alpha_1(0).\beta_1(0) & = & 0.009 x 0.74 & = & 0.00666 & & \\
\gamma_1(1) & = & \alpha_1(1).\beta_1(1) & = & 0.081 x 0.26 & = & \mathbf{0.02106} & & \text{*Most probable state is } S_1 \text{ at time } t_1 \\
\gamma_2(0) & = & \alpha_2(0).\beta_2(0) & = & 0.02116 x 1 & = & \mathbf{0.02116} & & \text{*Most probable state is } S_0 \text{ at time } t_2 \\
\gamma_2(1) & = & \alpha_2(1).\beta_2(1) & = & 0.00666 x 1 & = & 0.00666 & &
\end{array}
$$

So, most likely state sequence is $\mathbf{S_1, S_1, S_0}$ $\quad (\,or\ S_2, S_2, S_1)$

Method - 2: Viterbi Search

Initial probabilities;
$\pi(S_1) = 0.9$
$\pi(S_2) = 0.1$

For $t=0$
$S_1 \rightarrow S_1 \quad \Rightarrow \quad P(S_1).P(O=a \mid S_1).P(S_1 \mid S_1) = 0.072$
$S_2 \rightarrow S_1 \quad \Rightarrow \quad P(S_2).P(O=a \mid S_2).P(S_1 \mid S_2) = 0.018$
$max(0.072, 0.018) = 0.072$. So $S_1 \rightarrow S_1$ path is chosen

$S_2 \rightarrow S_1 \quad \Rightarrow \quad P(S_1).P(O=a \mid S_1).P(S_2 \mid S_1) = 0.018$
$S_2 \rightarrow S_2 \quad \Rightarrow \quad P(S_2).P(O=a \mid S_2).P(S_2 \mid S_2) = 0.072$
$max(0.072, 0.018) = 0.072$. So $S_2 \rightarrow S_2$ path is chosen

For $t=1$
$S_1 \rightarrow S_1 \quad \Rightarrow \quad 0.072.P(O=a \mid S_1).P(S_1 \mid S_1) = 0.00576$
$S_2 \rightarrow S_1 \quad \Rightarrow \quad 0.072.P(O=a \mid S_2).P(S_1 \mid S_2) = 0.01296$
$max(0.00576, 0.01296) = 0.01296$. So $S_2 \rightarrow S_1$ path is chosen

$S_2 \rightarrow S_1 \quad \Rightarrow \quad 0.072.P(O=a \mid S_1).P(S_2 \mid S_1) = 0.0144$
$S_2 \rightarrow S_2 \quad \Rightarrow \quad 0.072.P(O=a \mid S_2).P(S_2 \mid S_2) = 0.05184$
$max(0.0144, 0.05184) = 0.05184$. So $S_2 \rightarrow S_2$ path is chosen

For $t=2$
$S_1 \rightarrow b \quad \Rightarrow \quad 0.01296.P(O=b \mid S_1) = 0.011664$
$S_2 \rightarrow b \quad \Rightarrow \quad 0.05184.P(O=b \mid S_2) = 0.005184$
$max(0.011664, 0.005184) = 0.011664$. So $S_1 \rightarrow b$ emission is chosen

Finally, state sequence; $S_2 \rightarrow S_2 \rightarrow S_1$