CS540 Introduction to Artificial Intelligence
Lecture 23

Young Wu
Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

June 22, 2020

Midterm Format

Admin

Q1

(Monday)

- July 6 from 5 : 30 to 8 : 30
- A:
- B: I can make July 6
- C: I can only make July 7 ; 5:30 to 8:30
- D: I can not make July 6 or July 7
- E:

Midterm Reivew Session

5:30 - 8:30

- June 29 Dan will go through selected Homework questions and Past Exam questions, not recorded, notes will be posted.
- Dandi will go through the same questions this Thursday and Friday (June 18 and 19) 12 : 30 to 1 : 45 for section 1, you can use the guest link to attend too.

marginalization

$$\Pr\{w_3 = G, \underline{w_2 = I} \mid w_1 = I\} = 0.1 \cdot 0.7$$

$$+ \Pr\{w_3 = G, \underline{w_2 = am} \mid w_1 = I\} = 0.2 \cdot 0.5$$

$$+ \Pr[w_3 = G, \underline{w_2 = G} \mid w_1 = I] = 0.7 \cdot 0.3$$

Markov Chain Review

Quiz

- Given the transition matrix for "I", "am", "Groot", what is the probability that the third word is "Groot" given the first is "I"?

from to → next

current

$$
\begin{array}{c|c}
 & \begin{array}{ccc} I & am & Groot \end{array} \\
\begin{array}{c} I \\ am \\ Groot \end{array} & \begin{bmatrix} \boxed{0.1} & 0.2 & \boxed{0.7} \\ 0.2 & 0.3 & 0.5 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}
\end{array}
$$

Q2

$$\underbrace{Pr\{w_3 = Groot \mid w_1 = I\}}$$

$$Pr\{w_3 = Groot \mid w_2 = I,\ am,\ Groot\} \cdot Pr\{w_2 \mid w_1 = I\}$$

- A: 0.7
- B: $0.2 \cdot 0.4 + 0.3 \cdot 0.3$
- C: $0.2 \cdot 0.5 + 0.7 \cdot 0.3$
- $\boxed{\text{D}}$: $0.1 \cdot 0.7 + 0.2 \cdot 0.5 + 0.7 \cdot 0.3$
- E: $0.3 \cdot 0.3 + 0.2 \cdot 0.4 + 0.1 \cdot 0.3$

Causal Chain Review

Quiz

Q3

- Suppose the Bayesian Network is $A \to B \to C$, what is $\mathbb{P}\{A = 1, C = 1\}$?

CPT

$$\mathbb{P}\{A = 1\} = 0.4$$
$$\mathbb{P}\{B = 1|A = 1\} = 0.8, \mathbb{P}\{B = 1|A = 0\} = 0.1$$
$$\mathbb{P}\{C = 1|B = 1\} = 0.3, \mathbb{P}\{C = 1|B = 0\} = 0.7$$

$\Pr\{A, B, C\} + P\{A, \neg B, C\}$

$\Pr\{A=1\} \Pr\{B=1|A=1\} \Pr\{C=1|B=1\}$

$0.4 \cdot 0.8 \cdot 0.3 + 0.4 \cdot 0.2 \cdot 0.7$

- A: $0.4 \cdot 0.3$
- B: $0.4 \cdot 0.8 \cdot 0.3$
- C: $0.4 \cdot 0.8 \cdot 0.3 + 0.4 \cdot 0.2 \cdot 0.7$
- D: $0.4 \cdot 0.8 \cdot 0.3 + 0.4 \cdot 0.1 \cdot 0.7$
- E: $0.4 \cdot 0.8 \cdot 0.3 + 0.4 \cdot 0.2 \cdot 0.3$

$\Pr\{\neg B | A\}$

$1 - 0.8 = 0.2$

Causal Chain Review 2

Quiz

- $\mathbb{P}\{A = 1 | C = 1\} = \dfrac{\mathbb{P}\{A = 1, C = 1\}}{\mathbb{P}\{C = 1\}}$. What is $\mathbb{P}\{C = 1\}$?

Q4 (last)

$$\mathbb{P}\{A = 1\} = 0.4 \quad 0.6 \quad 1-0.4$$

$$\mathbb{P}\{B = 1 | A = 1\} = 0.8, \mathbb{P}\{B = 1 | A = 0\} = 0.1 \quad 1-0.1$$

$$\mathbb{P}\{C = 1 | B = 1\} = 0.3, \mathbb{P}\{C = 1 | B = 0\} = 0.7$$

$$\Pr(C, A, B) + \Pr\{C, \neg A, B\} + \Pr\{C, A, \neg B\} + \Pr\{C, \neg A, \neg B\}$$

- A: $0.3 \cdot 0.8 \cdot 0.4 + 0.3 \cdot 0.1 \cdot 0.6 + 0.7 \cdot 0.8 \cdot 0.4 + 0.7 \cdot 0.1 \cdot 0.6$
- B: $0.3 \cdot 0.8 \cdot 0.4 + 0.3 \cdot 0.1 \cdot 0.6 + 0.7 \cdot 0.2 \cdot 0.4 + 0.7 \cdot 0.9 \cdot 0.6$
- C: $0.3 \cdot 0.8 \cdot 0.4 + 0.3 \cdot 0.1 \cdot 0.4 + 0.7 \cdot 0.8 \cdot 0.4 + 0.7 \cdot 0.1 \cdot 0.4$
- D: $0.3 \cdot 0.8 \cdot 0.4 + 0.3 \cdot 0.1 \cdot 0.4 + 0.7 \cdot 0.2 \cdot 0.4 + 0.7 \cdot 0.9 \cdot 0.4$
- E: $0.3 \cdot 0.8 \cdot 0.4 + 0.3 \cdot 0.8 \cdot 0.6 + 0.7 \cdot 0.2 \cdot 0.4 + 0.7 \cdot 0.2 \cdot 0.6$

Causal Chain Review Derivation

Quiz

Special Bayesian Network for Sequences

Motivation

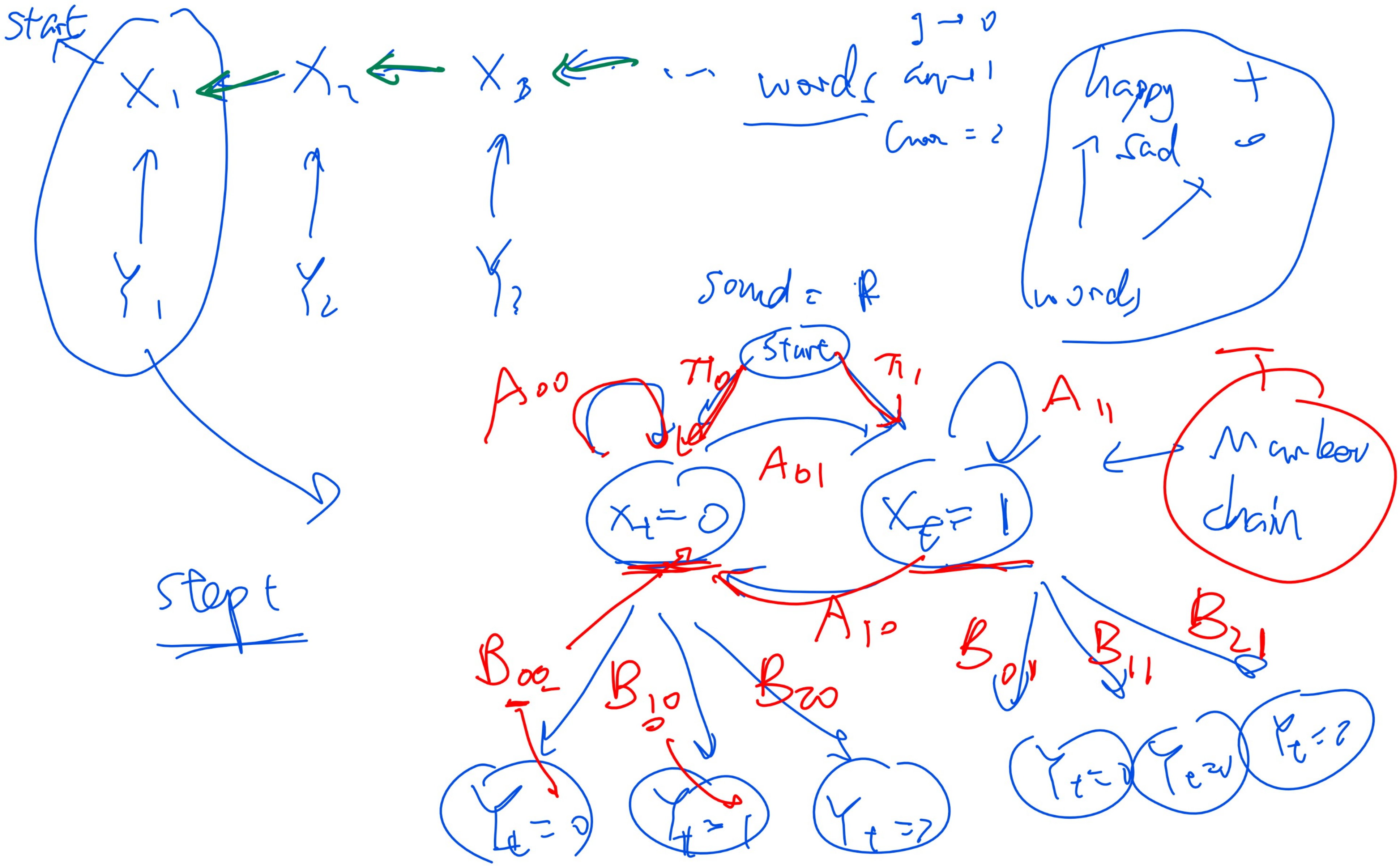
also estimate # possible X

iHMM

- A sequence of features $X_1, X_2, \ldots$ can be modeled by a Markov Chain but they are not observable.
- A sequence of labels $Y_1, Y_2, \ldots$ depends only on the current hidden features and they are observable.
- This type of Bayesian Network is called a Hidden Markov Model.

Hidden Markov Model Diagram

Motivation



Evaluation and Training

Motivation

- There are three main tasks associated with an HMM.

1. Evaluation problem: finding the probability of an observed sequence given an HMM: $y_1, y_2, \ldots$
2. Decoding problem: finding the most probable hidden sequence given the observed sequence: $x_1, x_2, \ldots$ missing
3. Learning problem: finding the most probable HMM given an observed sequence: $\pi, A, B, \ldots$

Evaluation Problem

Definition

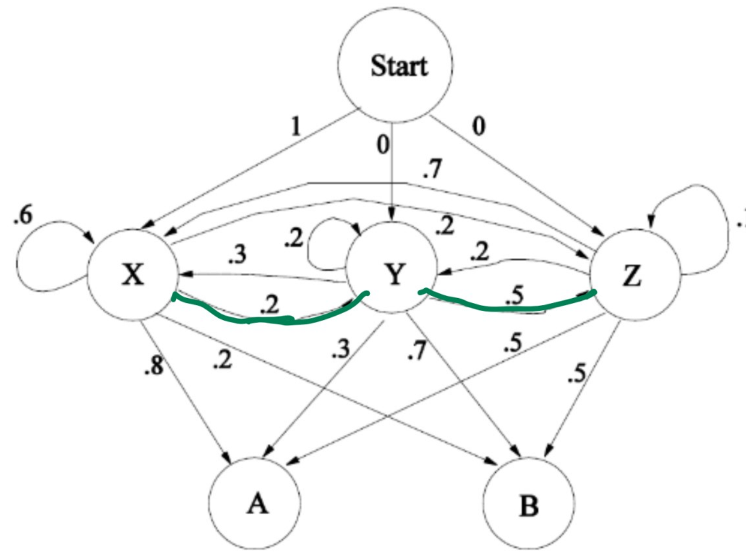
- The task is to find the probability $\mathbb{P}\{y_1, y_2, ..., y_T | \pi, A, B\}$.

$$\begin{aligned}
&\mathbb{P}\{y_1, y_2, ..., y_T | \pi, A, B\} \\
&= \sum_{x_1, x_2, ..., x_T} \mathbb{P}\{y_1, y_2, ..., y_T | x_1, x_2, ..., x_T\} \mathbb{P}\{x_1, x_2, ..., x_T\} \\
&= \sum_{x_1, x_2, ..., x_T} \left(\prod_{t=1}^{T} B_{y_t x_t} \right) \left(\pi_{x_1} \prod_{t=2}^{T} A_{x_{t-1} x_t} \right)
\end{aligned}$$

- This is also called the Forward Algorithm.

Evaluation Problem Example, Part 1

Definition



- Fall 2018 Final Q28 and Q29
- Compute $\mathbb{P}\left\{X_4 = Y, X_5 = Z | X_3 = X\right\}$.
- Compute $\mathbb{P}\left\{X_1 = X, X_2 = Z | Y_1 = A, Y_2 = B\right\}$.

$$Pr\{X_5 = Z \mid X_4 = Y\} \cdot Pr\{X_4 = Y \mid X_3 = X\}$$

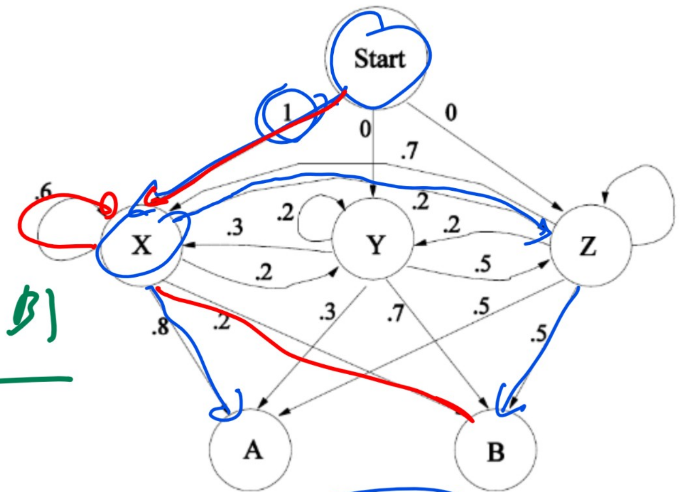
A_{YZ} 0.5 $\cdot$ $A_{XY} = 0.2$

0.1

Evaluation Problem Example, Part 2

Definition

$X_1 = X,\ X_2 = Z \mid Y_1 = A,\ Y_2 = B$



$$= \frac{\Pr\{X_1 = X,\ X_2 = Z,\ Y_1 = A,\ Y_2 = B\}}{\Pr\{Y_1 = A,\ Y_2 = B\}}$$

9

$X_1 \quad X_2$

X X

Y Y

X Z

6 impossible

$\Pr\{X_1 = X\} \cdot \Pr\{X_2 = Z \mid X_1 = X\}$

$\Pr\{Y_1 = A \mid X_1 = X\} \cdot \Pr\{Y_2 = B \mid X_2 = Z\}$

$= 1 \cdot 0.2 \cdot 0.8 \cdot 0.5$

$\Pr\{X_1 = X_2 = X,\ Y_1 = A,\ Y_2 = B\} = 1 \cdot 0.6 \cdot 0.8 \cdot 0.2$

Evaluation Problem Example, Part 3

Definition

Evaluation Problem Example, Part 4

Definition

Decoding Problem

Definition

- The task is to find $x_1, x_2, ..., x_T$ that maximizes $\mathbb{P}\{x_1, x_2, ..., x_T | y_1, y_2, ..., y_T, \pi, A, B\}$.
- Direct computation is too expensive.
- Dynamic programming needs to be used to save computation.
- This is called the Viterbi Algorithm. 577

Viterbi Algorithm Value Function

Definition

- Define the value functions to keep track of the maximum probabilities at each time t and for each state k.

$$
\begin{aligned}
V_{1,k} &= \mathbb{P}\left\{y_1 | X_1 = k\right\} \cdot \mathbb{P}\left\{X_1 = k\right\} \\
&= B_{y_1 k} \pi_k \\
V_{t,k} &= \max_x \mathbb{P}\left\{y_t | X_t = k\right\} \mathbb{P}\left\{X_t = k | X_{t-1} = x\right\} V_{1,k} \\
&= \max_x B_{y_t k} A_{kx} V_{1,k}
\end{aligned}
$$

Viterbi Algorithm Policy Function

Definition

- Define the policy functions to keep track of the x_t that maximizes the value function.

$$\text{policy}_{\,t,k} = \arg\max_x B_{y_t k} A_{kx} V_{1,k}$$

- Given the policy functions, the most probable hidden sequence can be found easily.

$$x_T = \arg\max_x V_{T,x}$$
$$x_t = \text{policy}_{\,t+1, x_{t+1}}$$

Dynamic Programming Diagram

Definition

Viterbi Algorithm Diagram

Definition

Expectation-Maximization Algorithm (for HMM), Part 1

Algorithm

- Initialize the hidden Markov model.

$$\pi \sim \mathrm{D}\left(|X|\right), A \sim \mathrm{D}\left(|X|, |X|\right), B \sim \mathrm{D}\left(|Y|, |X|\right)$$

- Perform the forward pass.

$$\alpha_{i,t} \text{ represents } \mathbb{P}\left\{y_1, y_2, ..., y_t, X_t = i | \pi, A, B\right\}$$
$$\alpha_{i,1} = \pi_i B_{y_1,i}$$
$$\alpha_{i,t+1} = \sum_{j=1}^{|X|} \alpha_{j,t} A_{ji} B_{y_{t+1} i}$$

Expectation-Maximization Algorithm (for HMM), Part 2

Algorithm

- Perform the backward pass.

$$\beta_{i,t} \text{ represents } \mathbb{P}\left\{y_{t+1}, y_{t+2}, ..., y_T | X_t = i, \pi, A, B\right\}$$

$$\beta_{i,T} = 1$$

$$\beta_{i,t} = \sum_{j=1}^{|X|} A_{ij} B_{y_{t+1} j} \beta_{j,t+1}$$

Expectation-Maximization Algorithm (for HMM), Part 3

Algorithm

- Define the conditional hidden state probabilities for each training sequence n.

$\approx$

$$\gamma_{n,i,t} = \text{ represents } \mathbb{P}\left\{X_t = i | y_1, y_2, ..., y_T, \pi, A, B\right\}$$

$$\gamma_{n,i,t} = \frac{\alpha_{i,t}\beta_{i,t}}{\sum_{j=1}^{|X|} \alpha_{j,t}\beta_{j,t}}$$

Come

frequency of $X_t = i$ in the training set

HMM

observe Y ←

not $\cancel{X}$ ←

Expectation-Maximization Algorithm (for HMM), Part 4

Algorithm

- Define the conditional hidden state probabilities for each training sequence n.

$\xi_{n,i,j,t}$ represents $\mathbb{P}\{X_t = i, X_{t+1} = j | y_1, y_2, ..., y_T, \pi, A, B\}$

$C_{w_{t-1} w_t}$

$$\xi_{n,i,j,t} = \dfrac{\alpha_{i,t} A_{ij} \beta_{j,t+1} B_{y_{t+1} j}}{\displaystyle\sum_{k=1}^{|X|} \sum_{l=1}^{|X|} \alpha_{k,t} A_{kl} \beta_{l,t+1} B_{y_{t+1} w}}$$

$\Pr\{ w_t | w_{t-1} \}$

Expectation-Maximization Algorithm (for HMM), Part 5

Algorithm

- Update the model.

$$\pi'_i = \frac{\sum_{n=1}^{N} \gamma_{n,i,1}}{N}$$

$$\frac{x_1 = i}{N}$$

$$A'_{ij} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T-1} \xi_{n,i,j,t}}{\sum_{n=1}^{N} \sum_{t=1}^{T-1} \gamma_{n,i,t}}$$

$$\frac{x_t = j,\ x_{t-1} = i}{x_{t-1} = i}$$

Expectation-Maximization Algorithm (for HMM), Part 6

Algorithm

@ 7:00

- Update the model, continued.

$$B'_{ij} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T} \mathbb{1}_{\{y_{n,t}=j\}} \gamma_{n,i,t}}{\sum_{n=1}^{N} \sum_{t=1}^{T} \gamma_{n,i,t}}$$

$$\frac{C \quad X_t = i,\ Y_t = j}{C \quad X_t = i}$$

- Repeat until π, A, B converge.

CPT