CMPSCI 240: "Reasoning Under Uncertainty"

Lecture 22

Prof. Hanna Wallach
wallach@cs.umass.edu

April 12, 2012

Reminders

- Check the course website: `http://www.cs.umass.edu/~wallach/courses/s12/cmpsci240/`
- Sixth homework is due TOMORROW

Grades

1. Add discussion section scores, divide by 14, multiply by 10
2. Add homework scores, divide by 250, multiply by 30
3. Divide midterm score by 100, multiply by 30

4. Add 1–3 to obtain your score (max. possible is 70)
5. If your score is 40 or less, you're in danger of getting a D

Recap

Last Time: Markov Chains

- There is some state space $\mathcal{S}$ that describes the system

Last Time: Markov Chains

- There is some state space $\mathcal{S}$ that describes the system
- State of the system at any time t is denoted X_t

Last Time: Markov Chains

- There is some state space $\mathcal{S}$ that describes the system
- State of the system at any time t is denoted X_t
- The system changes state according to transition probabilities

Last Time: Markov Chains

- There is some state space $\mathcal{S}$ that describes the system
- State of the system at any time t is denoted X_t
- The system changes state according to transition probabilities
- Probability of making a transition from state i to state j:

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

Last Time: Markov Chains

- There is some state space $\mathcal{S}$ that describes the system
- State of the system at any time t is denoted X_t

- The system changes state according to transition probabilities
- Probability of making a transition from state i to state j:

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

- Transition probabilities out of any state sum to one

Last Time: Markov Property

- Markov property: the probability of being in some state at time $t+1$ depends only on the state at time t:

$$P(X_{t+1}=j \mid X_t=i, \ldots, X_0=a) = P(X_{t+1}=j \mid X_t=i) = p_{ij}$$

Last Time: Markov Property

- Markov property: the probability of being in some state at time $t+1$ depends only on the state at time t:

$$P(X_{t+1}=j \mid X_t=i,\ldots,X_0=a) = P(X_{t+1}=j \mid X_t=i) = p_{ij}$$

- Can compute the probability of any sequence of states using the multiplication rule and the Markov property:

$$P(X_1=b, X_2=c, X_3=d, X_4=e \mid X_0=a) = p_{ab}\, p_{bc}\, p_{cd}\, p_{de}$$

Examples of Markov Chains

Examples of Markov Chains

- e.g., a kitten moves along a straight line in unit increments. At each time period, it moves one unit to the left with probability 0.3, one unit to the right with probability 0.3, and stays in place with probability 0.4, independent of the past history of movements. Two robots are lurking at positions 1 and m: if the kitten lands on either position, it is captured by a robot and consumed. Draw the state transition diagram for the case where $m = 4$. Write down the transition probabilities for the general case. What is the probability of the kitten moving right, right, no move, right, left, right, assuming that the kitten starts in position 2 and $m > 5$? What if $m = 5$?

Examples of Markov Chains

- e.g., consider a line at the Apple store. Every minute someone is served with probability $1/2$. Every minute someone joins the line with probability 1 if the line has length 0, with probability $2/3$ if the line has length 1, with probability $1/3$ if the line has length 2, and with probability 0 if the line has length 3.

Markov Text Generation

Markov Text Generation

- Each state is a word, e.g., $\mathcal{S} = \{\text{cat}, \text{dog}, \text{the}, \text{which}\}$

Markov Text Generation

- Each state is a word, e.g., $\mathcal{S} = \{\text{cat}, \text{dog}, \text{the}, \text{which}\}$
- When the system transitions into a state, the word corresponding to that state is "generated", i.e., printed

Markov Text Generation

- Each state is a word, e.g., $\mathcal{S} = \{\text{cat}, \text{dog}, \text{the}, \text{which}\}$
- When the system transitions into a state, the word corresponding to that state is "generated", i.e., printed
- The probability of generating some word at time t depends on the word that was generated at time $t-1$ (i.e., the state at time $t-1$) not on any previously-generated words

Building a Generator from a Document

- Create a state for every unique word + "START"

Building a Generator from a Document

- Create a state for every unique word + "START"
- Compute the transition probability for every pair of states:

$$P(X_{t+1}=j \mid X_t=i) = \frac{\#\text{ transitions to } j \text{ from } i}{\#\text{ transitions from } i}$$

Building a Generator from a Document

- Create a state for every unique word + "START"
- Compute the transition probability for every pair of states:

$$P(X_{t+1}=j \mid X_t=i) = \frac{\#\text{ transitions to } j \text{ from } i}{\#\text{ transitions from } i}$$

- If word j occurs immediately after word i then there will be a non-zero transition probability from word i to word j

Building a Generator from a Document

- Create a state for every unique word + “START”
- Compute the transition probability for every pair of states:

$$P(X_{t+1}=j \mid X_t=i) = \frac{\#\text{ transitions to } j \text{ from } i}{\#\text{ transitions from } i}$$

- If word j occurs immediately after word i then there will be a non-zero transition probability from word i to word j
- Will the generated text be identical to the document?

Examples of Computing Transition Probabilities

- e.g., compute the transition probabilities for “the dog which ate the cat which ate the dog which ate the cat”

Smoothing

- Only “observed” transitions have non-zero probability

Smoothing

- Only “observed” transitions have non-zero probability
- What if we want any transition to be possible?

Smoothing

- Only "observed" transitions have non-zero probability
- What if we want any transition to be possible?

- Pretend each transition (except for transitions to START) occurred one more time than it actually occurred:

$$P(X_{t+1}=j \mid X_t=i) = \frac{\#\text{ transitions to } j \text{ from } i + 1}{\#\text{ transitions from } i + \underbrace{|\mathcal{S}|-1}_{=W}}$$

Adding More Context

- Each state is a pair of words, e.g., $\mathcal{S} = \{\ldots, \text{the dog}, \ldots\}$
- Also add “START START” and, e.g., “START the”

Adding More Context

- Each state is a pair of words, e.g., $\mathcal{S} = \{\ldots, \text{the dog}, \ldots\}$
- Also add "START START" and, e.g., "START the"
- The second word is the word generated in that state

Adding More Context

- Each state is a pair of words, e.g., $\mathcal{S} = \{\ldots, \text{the dog}, \ldots\}$
- Also add "START START" and, e.g., "START the"
- The second word is the word generated in that state
- The first word is the word generated at the previous time step

Adding More Context

- Each state is a pair of words, e.g., $\mathcal{S} = \{\ldots, \text{the dog}, \ldots\}$
- Also add "START START" and, e.g., "START the"
- The second word is the word generated in that state
- The first word is the word generated at the previous time step
- The system can only transition between states that are consistent, e.g., from "the dog" to "dog ate"

More Context: Markov Property

- The Markov property holds: the probability of being in some state at time $t + 1$ depends only on the state at time t

More Context: Markov Property

- The Markov property holds: the probability of being in some state at time $t + 1$ depends only on the state at time t
- Because states are pairs of words, the word generated at time $t + 1$ depends on the words generated at times t and $t - 1$

More Context: Smoothing

- Need to add states/transitions corresponding to “unobserved” word pairs (except for “<?> START”): $W^2 + W + 1$ states

More Context: Smoothing

- Need to add states/transitions corresponding to “unobserved” word pairs (except for “<?> START”): $W^2 + W + 1$ states
- Smoothing doesn’t result in a fully-connected graph: inconsistent transitions must never occur

More Context: Smoothing

- Need to add states/transitions corresponding to "unobserved" word pairs (except for "<?> START"): $W^2 + W + 1$ states
- Smoothing doesn't result in a fully-connected graph: inconsistent transitions must never occur
- Smoothed transition probabilities are:

$$P(X_{t+1} = jk \mid X_t = ij) = \frac{\#\text{ transitions from } ij \text{ to } jk + 1}{\#\text{ transitions from } ij + W}$$

In the Real World

- Cell phones: predict the next word/letter in a SMS

In the Real World

- Cell phones: predict the next word/letter in a SMS
- Spam: generate English-like text to fool spam detectors

In the Real World

- Cell phones: predict the next word/letter in a SMS
- Spam: generate English-like text to fool spam detectors
- Can use a Markov text generator built from English text to score how "English-like" other documents are

In the Real World

- Cell phones: predict the next word/letter in a SMS
- Spam: generate English-like text to fool spam detectors
- Can use a Markov text generator built from English text to score how "English-like" other documents are
- `http://www.inference.phy.cam.ac.uk/dasher/`

For Next Time

- Check the course website: `http://www.cs.umass.edu/~wallach/courses/s12/cmpsci240/`
- Sixth homework is due TOMORROW