Hidden Markov Models

CISC 5800
Professor Daniel Leeds

Representing sequence data

- Spoken language
- DNA sequences
- Daily stock values

Example: spoken language

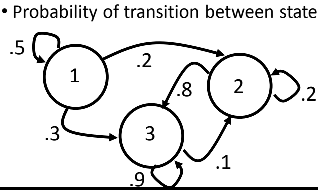
F?r plu? fi?e is nine

- Between F and r expect a vowel: “aw”, “ee”, “ah”; NOT “oh”, “uh”
- At end of “plu” expect consonant: “g”, “m”, “s”; NOT “d”, “p”

Markov Models

Start with:

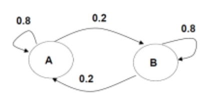
- *n* states: $s_1, \ldots, s_n$
- Probability of initial start states: $\Pi_1, \ldots, \Pi_n$
- Probability of transition between states: $A_{i,j} = P(q_t = s_i | q_{t-1} = s_j)$



A dice-y example

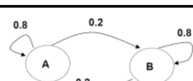
$\Pi_A = 0.3, \Pi_B = 0.7$

- Two colored die



- What is the probability we start at s_A?
- What is the probability we have the sequence of die choices: s_A, s_A?
- What is the probability we have the sequence of die choices: s_B, s_A, s_B, s_A?

A dice-y example



- What is the probability we have the sequence of die choices: s_B, s_A, s_B, s_A?

$\Pi_A = 0.3, \Pi_B = 0.7$

- Dynamic programming: find answer for q_t, then compute q_{t+1}

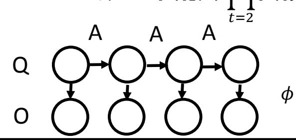
State\Time	t_1	t_2	t_3
s_A	0.3		
s_B	0.7		

$$p_t(i) = \sum_j p(q_t = s_i | q_{t-1} = s_j) p_{t-1}(j)$$

Hidden Markov Models

- Actual state q “hidden”
- State produces visible data o: $\phi_{i,j} = P(o_t = x_i | q_t = s_j)$
- Compute

$$P(\boldsymbol{O}, \boldsymbol{Q} | \boldsymbol{\theta}) = p(q_1 | \pi) \prod_{t=2}^{T} p(q_t | q_{t-1}, \boldsymbol{A}) \prod_{t=1}^{T} p(o_t | q_t, \boldsymbol{\phi})$$



Deducing die based on observed "emissions"

- Each color is biased
- A (red) B (blue)

o	$P(o\|s_A)$	$P(o\|s_B)$
1	.3	.1
2	.2	.1
3	.2	.1
4	.1	.2
5	.1	.2
6	.1	.3

- We see: 5 What is probability of o=5 | B (blue)
- We see: 5, 3 What is probability of **o**=5,3 | B, B?
  What is probability of **o**=5,3 and **s**=B,B
  What is MOST probable **s** | **o**=5,3?

Goal: calculate most likely states given observable data

$$\arg\max_Q P(Q \mid O) = \arg\max_Q \frac{P(O \mid Q)P(Q)}{P(O)}$$
$$= \arg\max_Q P(O \mid Q)P(Q)$$

- Define and use $\delta_t(i)$

$$\delta_t(i) = \max_{q_1 \dots q_{t-1}} p(q_1 \dots q_{t-1} \wedge q_t = s_i \wedge O_1 \dots O_t)$$

- Find the most likely path from q_1 to q_t that
  - $q_t=s_i$
  - Outputs are $o_1, \dots, o_t$

Viterbi algorithm: $\delta_t(i)$

- $\delta_1(i) = \Pi_i P(o_1|q_1 = s_i) = \Pi_i \phi_{1,i}$
- $\delta_t(i) = \max_j \delta_{t-1}(j)\, P(o_t|q_t = s_i) P\left(q_t = s_i | q_{t-1} = s_j\right) =$ $\max_j \boldsymbol{\delta_{t-1}(j)\, \phi_{t,i} A_{i,j}}$
- $P(Q^*|O)=\text{argmax}_Q P(Q|O)$

Parameters in HMM

- Initial probabilities: π_i
- Transition probabilities $A_{i,j}$
- Emission probabilities $\phi_{i,j}$

How do we learn these values?

First, assume we know the states

Learning HMM parameters: π_i

$\mathbf{x}^1$: A,B,A,A,B
$\mathbf{x}^2$: B,B,B,A,A
$\mathbf{x}^3$: A,A,B,A,B
⋮

Compute MLE for each parameter

$$\pi^* = \underset{\pi}{\text{argmax}} \prod_k \pi(q_1) \prod_{t=2}^{T} p(q_t|q_{t-1})$$

$$\pi_A = \frac{\#D(q_1 = s_A)}{\#D}$$

First, assume we know the states

Learning HMM parameters: $A_{i,j}$

$\mathbf{x}^1$: A,B,A,A,B
$\mathbf{x}^2$: B,B,B,A,A
$\mathbf{x}^3$: A,A,B,A,B
⋮

Compute MLE for each parameter

$$A^* = \underset{A}{\text{argmax}} \prod_k \pi(q_1) \prod_{t=2}^{T} p(q_t|q_{t-1})$$

$$A_{i,j} = \frac{\#D(q_t = s_i, q_{t-1} = s_j)}{\#D(q_{t-1} = s_j)}$$

First, assume we know the states

Learning HMM parameters: $\phi_{i,j}$

x[1]: A,B,A,A,B
o[1]: 2,5,3,3,6

x[2]: B,B,B,A,A
o[2]: 4,5,1,3,2

x[3]: A,A,B,A,B
o[3]: 1,4,5,2,6
⋮

Compute MLE for each parameter

$$\phi_{i,j} = \frac{\#D(o_t = i, q_t = s_j)}{\#D(q_t = s_j)}$$

13

Challenges in HMM learning

- Learning parameters (π, A, ϕ) with known states is not too hard
- BUT usually states are unknown
- If we had the parameters and the observations, we could figure out the states: Viterbi $P(Q^*|O)=\text{argmax}_Q P(Q|O)$

14

Expectation-Maximization, or "EM"

- Problem: Uncertain of y^i (class), uncertain of θ^i (parameters)
- Solution: Guess y^i, deduce θ^i, re-compute y^i, re-compute θ^i … etc.
  OR: Guess θ^i, deduce y^i, re-compute θ^i, re-compute y^i

Will converge to a solution

- E step: Fill in expected values for missing variables
- M step: Regular MLE given known and filled-in variables

Also useful when there are holes in your data

15

Computing states q_t

- Instead of picking one state: $q_t=s_i$, find $P(q_t=s_i|\mathbf{o})$

$$P(q_t = s_i | o_1, \cdots, o_T) = \frac{\alpha_t(i)\beta_t(i)}{\sum_j \alpha_t(j)\beta_t(j)}$$

Forward probability: $\boldsymbol{\alpha_t(i) = P(o_1 \dots o_t \wedge q_t = s_i)}$

Backward probability: $\boldsymbol{\beta_t(i) = P(o_{t+1} \dots o_T \wedge q_t = s_i)}$

16

Details of forward and backward probabilities

Forward probability: $\boldsymbol{\alpha_t(i) = P(o_1 \dots o_t \wedge q_t = s_i)}$

$$\alpha_1(i) = \phi_{o_1,i}\pi_i = P(o_1|q_1 = s_i)P(q_1 = s_i)$$

$$\alpha_t(i) = \phi_{o_t,i} \sum_j A_{i,j}\alpha_{t-1}(j)$$

$$\alpha_t(i) = P(o_t|q_t = s_i) \sum_j P(q_t = s_i | q_{t-1} = s_j)\alpha_{t-1}(j)$$

Backward probability: $\boldsymbol{\beta_t(i) = P(o_{t+1} \dots o_T \wedge q_t = s_i)}$

$$\beta_t(i) = \sum_j A_{j,i}\phi_{o_{t+1},j}\beta_{t+1}(j)$$

$$\beta_t(i) = \sum_j P(q_{t+1} = s_j | q_t = s_i)P(o_{t+1}|q_{t+1} = s_j)\beta_{t+1}(j)$$

Final $\boldsymbol{\beta}$: $\boldsymbol{\beta_{T-1}(i)}$

$$\beta_{T-1}(i) = \sum_j A_{j,i}\phi_{o_{T-1},j}$$

$$= P(q_T = s_j | q_{T-1} = s_i)P(o_T|q_T = s_j)$$

17

E-step: State probabilities

- One state:

$$P(q_t = s_i | o_1, \cdots, o_T) = \frac{\alpha_t(i)\beta_t(i)}{\sum_j \alpha_t(j)\beta_t(j)} = S_t(i)$$

- Two states in a row:

$$P(q_t = s_j, q_{t+1} = s_i | o_1, \cdots, o_T) = \frac{\alpha_t(j)A_{i,j}\phi_{o_{t+1},i}\beta_{t+1}(i)}{\sum_j \alpha_t(j)\beta_t(j)} = S_t(i,j)$$

18

Recall: when states known

- $\pi_A = \frac{\#D(q_1=s_A)}{\#D}$
- $A_{i,j} = \frac{\#D(q_t=s_i, q_{t-1}=s_j)}{\#D(q_{t-1}=s_j)}$
- $\phi_{i,j} = \frac{\#D(o_t=i)}{\#D(q_t=s_j)}$

19

M-step

- $A_{i,j} = \frac{\sum_t S_t(i,j)}{\sum_k \sum_t S_t(k,j)}$
- $\phi_{obs,i} = \frac{\sum_{t|o_t=obs} S_t(i)}{\sum_j \sum_{t|o_t=j} S_t(i)}$
- $\pi_i = \frac{\sum_{seq} S_1(i)}{\sum_j \sum_{seq} S_1(j)}$

Known states:

- $\pi_A = \frac{\#D(q_1=s_A)}{\#D}$
- $A_{i,j} = \frac{\#D(q_t=s_i, q_{t-1}=s_j)}{\#D(q_{t-1}=s_j)}$
- $\phi_{i,j} = \frac{\#D(o_t=i)}{\#D(q_t=s_j)}$

20

Review of HMMs in action

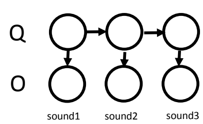
For classification, find highest probability class given features

Features for one sound:

- $[q_1, o_1, q_2, o_2, \ldots, q_T, o_T]$

Conclude word:

Generates states:



21