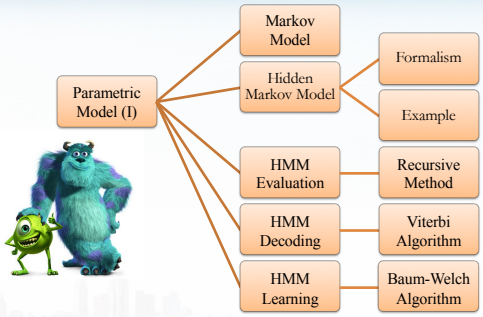
Lecture Notes on

Pattern Recognition

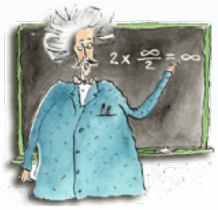
Session 04(A): Parametric Model (I)

Gang Li
School of Information Technology
Deakin University, VIC 3125, Australia

Road map



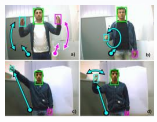
Markov Model



- First Order Markov Model
- Limitations

Sequential and Temporal Patterns

- Some applications have an inherent temporality.
  - Speech recognition
  - Gesture recognition
  - Human activity recognition



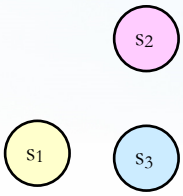
Sequential and Temporal Patterns

- ***Temporal*** patterns:
  - The order of data points is important
    - (i.e., time series).
  - Data can be represented by a number of states.
  - States at time t are influenced directly by states in previous time steps
    - (i.e., correlated).
- ***Sequential*** patterns:
  - The order of data points is irrelevant.

First Order Markov Model

Markov Model

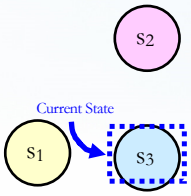
- States: $s_1, s_2 .. s_N$
- Time-steps: $t=0,1, \ldots$



First Order Markov Model

Markov Model

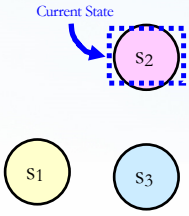
- States: $s_1, s_2 .. s_N$
- Time-steps: $t=0,1, \ldots$
- On the t-th time-step the system is in exactly one of the available states q_t



First Order Markov Model

Markov Model

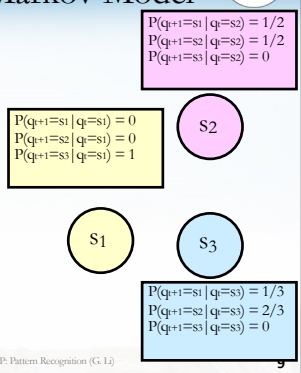
- States: $s_1, s_2 .. s_N$
- Time-steps: $t=0,1, \ldots$
- On the t-th time-step the system is in exactly one of the available states q_t
- The next state is chosen randomly by a distribution based on current status.



First Order Markov Model

Markov Model

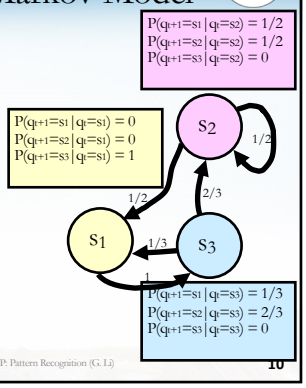
- States: $s_1, s_2 .. s_N$
- Time-steps: $t=0,1, \ldots$
- On the t-th time-step the system is in exactly one of the available states q_t
- The next state is chosen randomly by a distribution based on current status.



First Order Markov Model

Markov Model

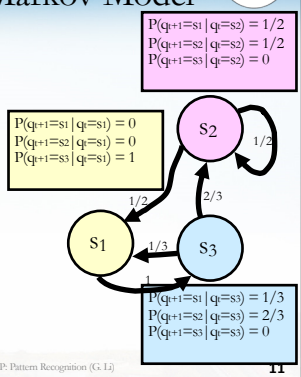
- States: $s_1, s_2 .. s_N$
- Time-steps: $t=0,1, \ldots$
- On the t-th time-step the system is in exactly one of the available states q_t
- The next state is chosen randomly by a distribution based on current status.



First Order Markov Model

Markov Property

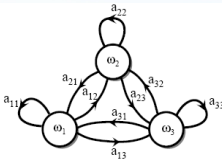
- q_{t+1} is conditionally independent of $\{ q_{t-1}, q_{t-2}, \ldots q_1, q_0 \}$ given q_t.
- In other words:
  - $P(q_{t+1} = s_j \mid q_t = s_i) = P(q_{t+1} = s_j \mid q_t = s_i$,any earlier history)
  - $a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i)$



First Order Markov Model

Markov Property

- q_{t+1} is conditionally independent of $\{ q_{t-1}, q_{t-2}, \ldots q_1, q_0 \}$ given q_t.
- In other words:
  - $P(q_{t+1} = s_j \mid q_t = s_i) = P(q_{t+1} = s_j \mid q_t = s_i$,any earlier history)
  - $a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i)$

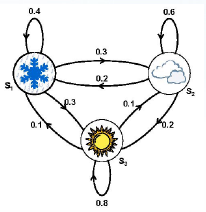


First Order Markov Model (Weather Prediction Model)

- Assume three weather states:
  - S_1: *Precipitation* (rain, snow, hail, etc.)
  - S_2: *Cloudy*
  - S_3: *Sunny*
- Transition Matrix

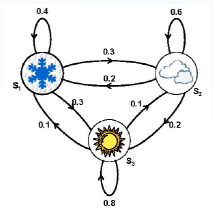
	S_1	S_2	S_3
S_1	0.4	0.3	0.3
S_2	0.2	0.6	0.2
S_3	0.1	0.1	0.8

$$a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i)$$



First Order Markov Model (Weather Prediction Model)

- The probability of a sequence of states $q^T=(q_1, q_2, \ldots, q_T)$ is the product of the corresponding transition probabilities
  - "**sunny-sunny-sunny-rainy-rainy-sunny-cloudy-sunny**" ?
  - $p(S_3)a_{33}a_{33}a_{31}a_{11}a_{13}a_{32}a_{23}$

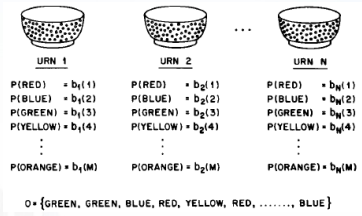


Limitations of Markov models

- In Markov models, each state is ***uniquely*** associated with an ***observable*** event.
  - Once an observation is made, the state of the system is trivially retrieved.
  - Such systems are ***not*** of practical use for most applications.

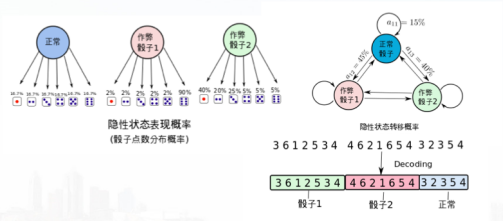
Limitations of Markov models

- From the sequence of balls you observed, figure out the sequence of URNs the balls are taken from?



Limitations of Markov models

- Given a sequence of dice rolls, assuming that we know that the casino player may switch between a fair and two loaded die:



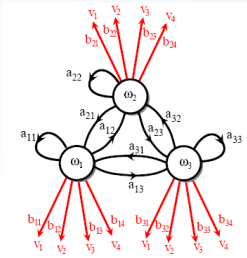
Hidden Markov Model

- HMM Formalism
- HMM Example

Hidden Markov Models

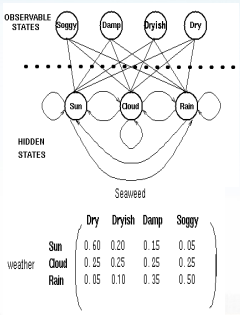
First Order HMM

- Augment Markov model such that when it is in a ***hidden*** state ω(t) it also emits some ***visible*** symbol v(t) among a set of possible symbols.
- We have access to the **visible states** only, while **hidden states** are unobservable.



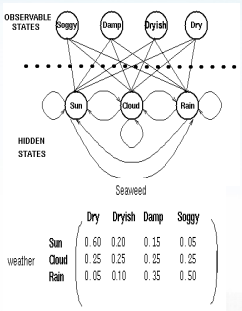
Hidden Markov Models

- **Hidden states** : the (TRUE) states of a system that may be described by a Markov process
  - e.g., the weather
- **Observable states** : the states of the process that are `visible'
  - e.g., seaweed dampness



Hidden Markov Models

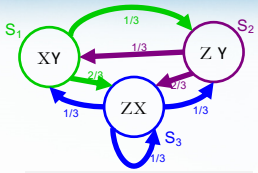
- **Initial Distribution** : contains the probability of the (hidden) model being in a particular initial hidden state.
- **State transition matrix** : holding the probability of a hidden state given the previous hidden state.
- **Output matrix** : containing the probability of observing a particular observable state given that the hidden model is in a particular hidden state.



HMM Formalism

- An HMM is defined by $\lambda=\{\Omega, V, P, A, B\}$:
  - Ω : $\{S_1 \ldots S_n\}$ are n possible ***states***
  - V : $\{O_1 \ldots O_m\}$ are m possible ***observations***
  - $P = \{\pi_1 \ldots \pi_n\}$ are the ***prior state probabilities***
  - $A = \{a_{ij}\}$ is the (n x n) ***state transition matrix***
  - $B = \{b_{ik}\}$ is the (n x m) ***observation state matrix***

An HMM Example



- Start randomly in state S_1 or S_2
- Choose one of the output symbols in each state at random.

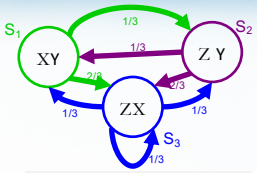
n = 3, m = 3

$\pi_1 = 1/2 \quad \pi_2 = 1/2 \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

$b_1(X) = 1/2 \quad b_1(Y) = 1/2 \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = 1/2 \quad b_2(Z) = 1/2$
$b_3(X) = 1/2 \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

An HMM Example



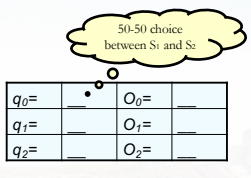
- Start randomly in state S_1 or S_2
- Choose one of the output symbols in each state at random.
- Let us generate a sequence of observations

n = 3, m = 3

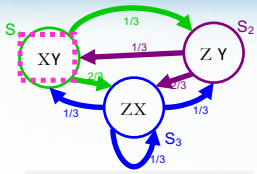
$\pi_1 = 1/2 \quad \pi_2 = 1/2 \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

$b_1(X) = 1/2 \quad b_1(Y) = 1/2 \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = 1/2 \quad b_2(Z) = 1/2$
$b_3(X) = 1/2 \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$



An HMM Example



$n = 3,\ m = 3$

$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

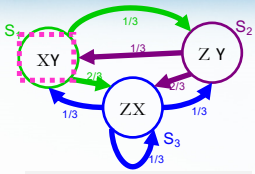
$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

- Start randomly in state S_1 or S_2
- Choose one of the output symbols in each state at random.
- Let us generate a sequence of observations

50-50 choice between X and Y

q_0=	S_1	O_0=	___
q_1=	___	O_1=	___
q_2=	___	O_2=	___

An HMM Example



$n = 3,\ m = 3$

$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

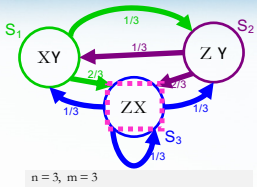
$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

- Start randomly in state S_1 or S_2
- Choose one of the output symbols in each state at random.
- Let us generate a sequence of observations

Goto S3 with probability 2/3 or S2 with prob. 1/3

q_0=	S_1	O_0=	X
q_1=	___	O_1=	___
q_2=	___	O_2=	___

An HMM Example



$n = 3,\ m = 3$

$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

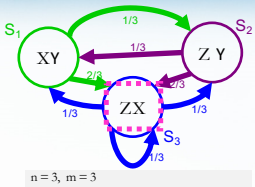
$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

- Start randomly in state S_1 or S_2
- Choose one of the output symbols in each state at random.
- Let us generate a sequence of observations

50-50 choice between X and Z

q_0=	S_1	O_0=	X
q_1=	S_3	O_1=	•
q_2=	___	O_2=	___

An HMM Example



$n = 3,\ m = 3$

$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

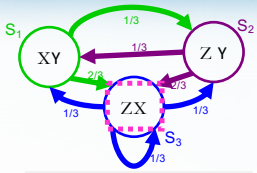
$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

- Start randomly in state S_1 or S_2
- Choose one of the output symbols in each state at random.
- Let us generate a sequence of observations

Each of the three next states is equally likely

q_0=	S_1	O_0=	X
q_1=	S_3	O_1=	X
q_2=	___	O_2=	___

An HMM Example



$n = 3,\ m = 3$

$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

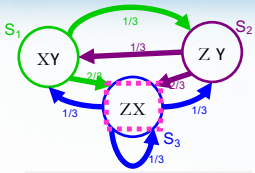
$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

- Start randomly in state S_1 or S_2
- Choose one of the output symbols in each state at random.
- Let us generate a sequence of observations

50-50 choice between X and Z

q_0=	S_1	O_0=	X
q_1=	S_3	O_1=	X
q_2=	S_3	O_2=	•

An HMM Example



$n = 3,\ m = 3$

$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

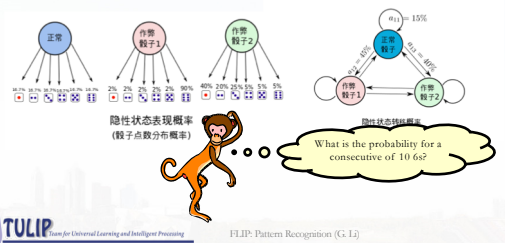
- Start randomly in state S_1 or S_2
- Choose one of the output symbols in each state at random.
- Let us generate a sequence of observations

This is what the observer has to work with…

q_0=	S_1	O_0=	X
q_1=	S_3	O_1=	X
q_2=	S_3	O_2=	Z

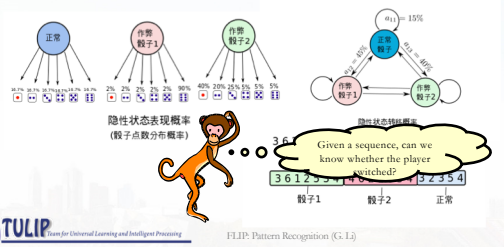
Another HMM Example

- Given a sequence of dice rolls, assuming that we know that the casino player may switch between a fair and two loaded die:



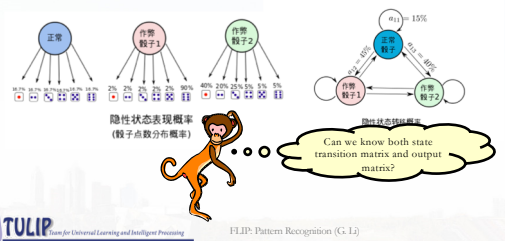
Another HMM Example

- Given a sequence of dice rolls, assuming that we know that the casino player may switch between a fair and two loaded die:



Another HMM Example

- Given a sequence of dice rolls, assuming that we know that the casino player may switch between a fair and two loaded die:



Three basic HMM problems

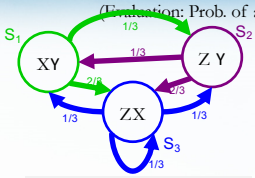
- **Evaluation**
  - Determine the probability that a particular sequence of ***visible states*** was generated by a given model
    - i.e., Forward/Backward algorithm
- **Decoding**
  - Given a sequence of ***visible states***, determine the most likely sequence of ***hidden states*** that led to those observations
    - i.e., using Viterbi algorithm
- **Learning**
  - Given a set of visible observations, determine a_{ij} and b_{jk}
    - i.e., using EM algorithm - Baum-Welch algorithm

HMM Evaluation

- HMM Evaluation
- Recursive Method

An HMM Example

(Evaluation: Prob. of a sequence of observations)



n = 3, m = 3

$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

- What is $P(\mathbf{O}) = P(O_1\ O_2\ O_3) = P(O_1 = X \wedge O_2 = X \wedge O_3 = Z)$?
- Slow, stupid way:

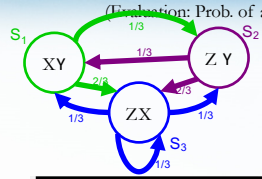
$$P(\mathbf{O}) = \sum_{\mathbf{Q} \in \text{Paths of length 3}} P(\mathbf{O} \wedge \mathbf{Q})$$

$$= \sum_{\mathbf{Q} \in \text{Paths of length 3}} P(\mathbf{O} \mid \mathbf{Q}) P(\mathbf{Q})$$

  - How do we compute P(Q) for an arbitrary path Q?
  - How do we compute P(O|Q) for an arbitrary path Q?

An HMM Example

(Evaluation: Prob. of a sequence of observations)



- What is $P(\mathbf{O}) = P(O_1\, O_2\, O_3) = P(O_1 = X \wedge O_2 = X \wedge O_3 = Z)$?
- Slow, stupid way:

$$P(\mathbf{O}) = \sum_{\mathbf{Q}\in\text{Paths of length 3}} P(\mathbf{O}\wedge\mathbf{Q})$$

$$= \sum_{\mathbf{Q}\in\text{Paths of length 3}} P(\mathbf{O}|\mathbf{Q})P(\mathbf{Q})$$

$P(Q) = P(q_1,q_2,q_3)$

$= P(q_1)\, P(q_2,q_3|q_1)$ (chain rule)

$= P(q_1)\, P(q_2|q_1)\, P(q_3|\, q_2,q_1)$ (chain)

$= P(q_1)\, P(q_2|q_1)\, P(q_3|\, q_2)$ (why?)

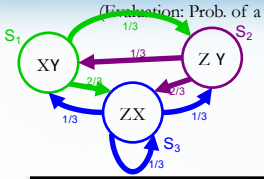
Example in the case $Q = S_1 S_3 S_3$:
$= 1/2 * 2/3 * 1/3 = 1/9$

How do we compute P(Q) for an arbitrary path Q?

- How do we compute $P(O|Q)$ for an arbitrary path Q?

An HMM Example

(Evaluation: Prob. of a sequence of observations)



- What is $P(\mathbf{O}) = P(O_1\, O_2\, O_3) = P(O_1 = X \wedge O_2 = X \wedge O_3 = Z)$?
- Slow, stupid way:

$$P(\mathbf{O}) = \sum_{\mathbf{Q}\in\text{Paths of length 3}} P(\mathbf{O}\wedge\mathbf{Q})$$

$$= \sum_{\mathbf{Q}\in\text{Paths of length 3}} P(\mathbf{O}|\mathbf{Q})P(\mathbf{Q})$$

$P(O|Q)$
$= P(O_1\, O_2\, O_3\, |q_1\, q_2\, q_3)$
$= P(O_1 \mid q_1)\, P(O_2 \mid q_2)\, P(O_3 \mid q_3)$
(why?)

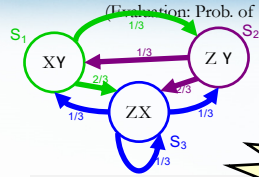
Example in the case $Q = S_1 S_3 S_3$:
$= P(X|\, S_1)\, P(X|\, S_3)\, P(Z|\, S_3) =$
$= 1/2 * 1/2 * 1/2 = 1/8$

- How do we compute P(Q) for an arbitrary path Q?

How do we compute P(O|Q) for an arbitrary path Q?

An HMM Example

(Evaluation: Prob. of a sequence of observations)



- What is $P(\mathbf{O}) = P(O_1\, O_2\, O_3) = P(O_1 = X \wedge O_2 = X \wedge O_3 = Z)$?

$n = 3,\ m = 3$

$\pi_1 = 1/2 \quad \pi_2 = 1/2 \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

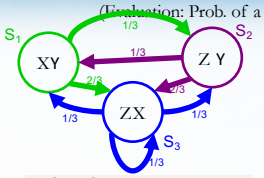
$b_1(X) = 1/2 \quad b_1(Y) = 1/2 \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = 1/2 \quad b_2(Z) = 1/2$
$b_3(X) = 1/2 \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

$P(\mathbf{O})$ would need 27 $P(Q)$ computations and 27 $P(O|Q)$ computations

A sequence of 20 observations would need $3^{20} =$ 3.5 billion computations and 3.5 billion $P(O|Q)$ computations

An HMM Example

(Evaluation: Prob. of a sequence of observations)



- Given observations $O_1\, O_2 \ldots O_T$, define

$\alpha_t(i) = P(O_1\, O_2 \ldots O_t \wedge q_t = S_i \mid \lambda)$

where $1 \le t \le T$

  - $\alpha_t(i)$ is the probability that: *in a random trial, we have seen the first **t** observations, and ended up in S_i as the **t**-th state visited.*
  - What is $\alpha_2(3)$?

$n = 3,\ m = 3$

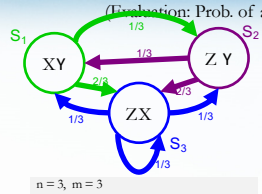
$\pi_1 = 1/2 \quad \pi_2 = 1/2 \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

$b_1(X) = 1/2 \quad b_1(Y) = 1/2 \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = 1/2 \quad b_2(Z) = 1/2$
$b_3(X) = 1/2 \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

An HMM Example

(Evaluation: Prob. of a sequence of observations)



- Given observations $O_1\, O_2 \ldots O_T$, define

$\alpha_t(i) = P(O_1\, O_2 \ldots O_t \wedge q_t = S_i \mid \lambda)$

where $1 \le t \le T$

$n = 3,\ m = 3$

$\pi_1 = 1/2 \quad \pi_2 = 1/2 \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

$b_1(X) = 1/2 \quad b_1(Y) = 1/2 \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = 1/2 \quad b_2(Z) = 1/2$
$b_3(X) = 1/2 \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

$\alpha_t(i) = P(O_t \wedge q_t = S_i)$
$= P(q_t = S_i) P(O_t | q_t = S_i)$

$\alpha_{t+1}(j) = P(O_1 O_2 \ldots O_t O_{t+1} \wedge q_{t+1} = S_j)$

$= \sum_i P(O_1 O_2 \ldots O_t \wedge q_t = S_i \wedge O_{t+1} \wedge q_{t+1} = S_j)$

$= \sum_i P(O_1 O_2 \ldots O_t \wedge q_t = S_i) P(O_{t+1}, q_{t+1} = S_j | O_1 O_2 \ldots O_t \wedge q_t = S_i) P(O_1 O_2 \ldots O_t \wedge q_t = S_i)$

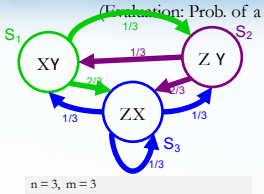
$= \sum_i P(O_{t+1}, q_{t+1} = S_j | q_t = S_i) \alpha_t(i)$

$= \sum_i P(q_{t+1} = S_j | q_t = S_i) P(O_{t+1} | q_{t+1} = S_j) \alpha_t(i)$

$= \sum_i a_{ij} b_j(O_{t+1}) \alpha_t(i)$

An HMM Example

(Evaluation: Prob. of a sequence of observations)



- Given observations $O_1 O_2 O_3 = XXZ$

$\alpha_t(i) = P(O_1 O_2 .. O_t \wedge q_t = S_i | \lambda)$

$\alpha_1(i) = b_i(O_1)\pi_i$

$\alpha_{t+1}(j) = \sum_i a_{ij} b_j(O_{t+1}) \alpha_t(i)$

$n = 3,\ m = 3$

$\pi_1 = 1/2 \quad \pi_2 = 1/2 \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

$b_1(X) = 1/2 \quad b_1(Y) = 1/2 \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = 1/2 \quad b_2(Z) = 1/2$
$b_3(X) = 1/2 \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

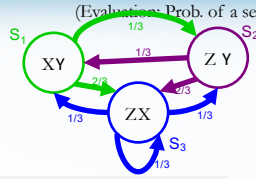
$\alpha_1(1) = \frac{1}{4} \quad \alpha_1(2) = 0 \quad \alpha_1(3) = 0$

$\alpha_2(1) = 0 \quad \alpha_2(2) = 0 \quad \alpha_2(3) = \frac{1}{12}$

$\alpha_3(1) = 0 \quad \alpha_3(2) = \frac{1}{72} \quad \alpha_3(3) = \frac{1}{72}$

An HMM Example

(Evaluation: Prob. of a sequence of observations)



n = 3, m = 3
$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

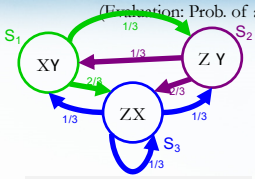
$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

- What is $P(\mathbf{O}) = P(O_1\, O_2\, O_3)$ $= P(O_1 = X \wedge O_2 = X \wedge O_3 = Z)$?
  - Better way?
  - **Exercise NOW**
- What is $P(q_3=S_2 \mid O_1\, O_2\, O_3)$?
  - **Exercise NOW**

$$\alpha_1(1)=\frac{1}{4} \quad \alpha_1(2)=0 \quad \alpha_1(3)=0$$
$$\alpha_2(1)=0 \quad \alpha_2(2)=0 \quad \alpha_2(3)=\frac{1}{12}$$
$$\alpha_3(1)=0 \quad \alpha_3(2)=\frac{1}{72} \quad \alpha_3(3)=\frac{1}{72}$$

An HMM Example

(Evaluation: Prob. of a sequence of observations)



n = 3, m = 3
$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

- We can cheaply compute $\alpha_t(i)=P(O_1O_2\ldots O_t \wedge q_t=S_i)$
- How can we cheaply compute $P(O_1O_2\ldots O_t)$?

$$\sum_{i=1}^{N}\alpha_t(i)$$

- How can we cheaply compute $P(q_t=S_i \mid O_1O_2\ldots O_t)$?

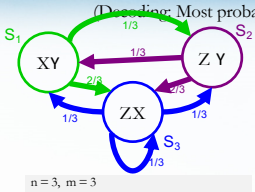
$$\frac{\alpha_t(i)}{\sum_{j=1}^{N}\alpha_t(j)}$$

HMM Decoding

- HMM Decoding
- Viterbi Algorithm

Viterbi Algorithm

(Decoding: Most probable path given observations)



n = 3, m = 3
$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

- What are the most probable path given observations $O_1\, O_2\, O_3$?

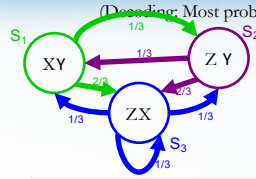
$$\underset{Q}{\text{argmax}}\ P(Q|O_1O_2\ldots O_T)$$

This is what the observer has to work with…

$q_0=$	S_1	$O_0=$	X
$q_1=$	S_3	$O_1=$	X
$q_2=$	S_3	$O_2=$	Z

Viterbi Algorithm

(Decoding: Most probable path given observations)



n = 3, m = 3
$\pi_1 = 1/2 \quad \pi_2 = ½ \quad \pi_3 = 0$

$a_{11} = 0 \quad a_{12} = 1/3 \quad a_{13} = 2/3$
$a_{21} = 1/3 \quad a_{22} = 0 \quad a_{23} = 2/3$
$a_{31} = 1/3 \quad a_{32} = 1/3 \quad a_{33} = 1/3$

$b_1(X) = ½ \quad b_1(Y) = ½ \quad b_1(Z) = 0$
$b_2(X) = 0 \quad b_2(Y) = ½ \quad b_2(Z) = 1/2$
$b_3(X) = ½ \quad b_3(Y) = 0 \quad b_3(Z) = 1/2$

- What are the most probable path given observations $O_1\, O_2\, O_3$?

$$\underset{Q}{\text{argmax}}\ P(Q|O_1O_2\ldots O_T)$$

Slow, stupid answer:

$$\underset{Q}{\text{argmax}}\ P(Q|O_1O_2\ldots O_T)$$
$$= \underset{Q}{\text{argmax}}\ \frac{P(O_1O_2\ldots O_T|Q)P(Q)}{P(O_1O_2\ldots O_T)}$$
$$= \underset{Q}{\text{argmax}}\ P(O_1O_2\ldots O_T|Q)P(Q)$$

Viterbi Algorithm

(Decoding: Most probable path given observations)

- We're going to compute the following variables:

$\delta_t(i) = \max_{q_1 q_2 \ldots q_{t-1}} P(q_1\, q_2 \ldots q_{t-1} \wedge q_t = S_i \wedge O_1 \ldots O_t)$

= Probability of Length **t-1** path with the maximum chance of:
…OCCURING
and
…ENDING UP IN STATE S_i
and
…PRODUCING OUTPUT $O_1 \ldots O_t$

- DEFINE: $\mathbf{mpp_t(i)}$ = that path,
  - So $\delta_t(i) = P(\mathbf{mpp_t(i)})$

Viterbi Algorithm

(Decoding: Most probable path given observations)

$$\delta_t(i) = \max_{q_1 q_2 \ldots q_{t-1}} P(q_1 q_2 \ldots q_{t-1} \wedge q_t = S_i \wedge O_1 O_2 .. O_t)$$

$$mpp_t(i) = \arg\max_{q_1 q_2 \ldots q_{t-1}} P(q_1 q_2 \ldots q_{t-1} \wedge q_t = S_i \wedge O_1 O_2 .. O_t)$$

$$\delta_1(i) = \max_{\text{one choice}} P(q_1 = S_i \wedge O_1)$$
$$= P(q_1 = S_i) P(O_1 | q_1 = S_i) = \pi_i b_i(O_1)$$

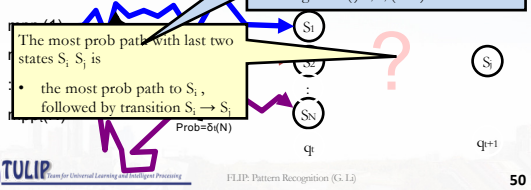
Viterbi Algorithm

(Decoding: Most probable path given observations)

- Suppose we have $\delta_t(i)$ and $mpp_t(i)$ for all i.
- Task:
  - How to get $\delta_{t+1}(j)$ and mpp

What is the prob of that path?
$\delta_t(i) x P(S_i \to S_j \wedge O_{t+1} | \lambda) = \delta_t(i)\, a_{ij}\, b_j(O_{t+1})$
- SO The most probable path to S_j has S_{i^*} as its penultimate state with $i^* = \text{argmax}\, \delta_t(i)\, a_{ij}\, b_j(O_{t+1})$

The most prob path with last two states S_i S_j is
- the most prob path to S_i, followed by transition $S_i \to S_j$

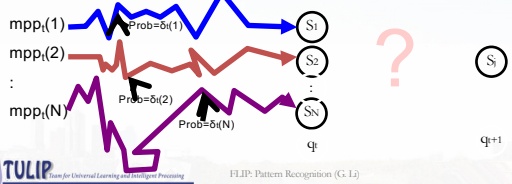


Viterbi Algorithm

(Decoding: Most probable path given observations)

- Suppose we have $\delta_t(i)$ and $mpp_t(i)$ for all i.
- Task:
  - How to get $\delta_{t+1}(j)$ and $mpp_{t+1}(j)$?
    - $\delta_{t+1}(j) = \delta_t(i^*)\, a_{ij}\, b_j(O_{t+1})$
    - $mpp_{t+1}(j) = mpp_{t+1}(i^*) S_{i^*}$

$i^* = \text{argmax}\, \delta_t(i)\, a_{ij}\, b_j(O_{t+1})$



Andrew Viterbi

"The algorithm was originally created for improving communication from space by being able to operate with a weak signal but today it has a multitude of applications," Viterbi

- Life Fellow Andrew Viterbi created an algorithm used in every cell phone and cofound Qualcomm.
- For the algorithm, which carries his name, he was awarded the Benjamin Franklin Medal in electrical engineering by the Franklin Institute in Philadelphia, one of the United States' oldest centres of science education and development.
- Another successful venture for the company was the creation of code-division multiple access (CDMA), which was introduced commercially in 1995 in cellphones and is still big today.

HMM Learning

- Learning HMM
- BAUM-WELCH algorithm

Learning an HMM

- We have been doing things like

$$P(O_1\, O_2 \ldots O_T \mid \lambda)$$

λ is the HMM parameters.

- But how do you know this λ?
  - Hand crafted? OR
  - But usually, especially in Speech or Genetics, it is better to infer it from large amounts of data
    - $O_1\, O_2 \ldots O_T$ with a big "**T**".

Learning an HMM

- Task: we have some observations $O_1\ O_2\ ..\ O_T$ and we want to estimate λ from them.
- Two approaches
  - MLE

    $\lambda = \text{argmax}\ P(O_1\ ..\ O_T \mid \lambda)$
  - B.E.
    - Work out $P(\lambda \mid O_1\ ..\ O_T)$, and then
      - take E[λ] or
      - max $P(\lambda \mid O_1\ ..\ O_T)$

MLE HMM from Observations

- **Expectation**

$\gamma_t(i) = P(q_t = S_i \mid O_1 O_2 .. O_T, \lambda)$

$\varepsilon_t(i,j) = P(q_t = S_i \wedge q_{t+1} = S_j \mid O_1 O_2 .. O_T, \lambda)$

$\sum_{t=1}^{T-1} \gamma_t(i)$ = expected number of transitions out of state i during path

$\sum_{t=1}^{T-1} \varepsilon_t(i,j)$ = expected number of transitions out of i and into j during path

MLE HMM from Observations

- **Maximization**

Notice that $\dfrac{\sum_{t=1}^{T-1} \varepsilon_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} = \dfrac{\begin{pmatrix}\text{expected frequency} \\ i \rightarrow j\end{pmatrix}}{\begin{pmatrix}\text{expected frequency} \\ i\end{pmatrix}}$

= Estimate of Prob(Next state S_j | This state S_i)

We can re-estimate

$a_{ij} \leftarrow \dfrac{\sum \varepsilon_t(i,j)}{\sum \gamma_t(i)}$

We can also re-estimate

$b_j(O_k) \leftarrow \dfrac{E[\#\ times\ it\ emits\ symbol\ k\ while\ at\ state\ \omega_j]}{E[\#\ times\ it\ emits\ any\ other\ symbol\ while\ at\ state\ \omega_j]}$

EM for HMMs (BAUM-WELCH algorithm)

Expectation

- If we knew λ we could estimate ***expectations*** such as
  - Expected number of times in state i
  - Expected number of transitions $i \rightarrow j$

Maximization

- If we knew quantities such as
  - Expected number of times in state i
  - Expected number of transitions $i \rightarrow j$
- We can MLE:
  $\lambda = \langle \{a_{ij}\}, \{b_i(j)\}, \pi_i \rangle$

EM for HMMs (BAUM-WELCH algorithm)

- **BAUM-WELCH algorithm**

1. Get your observations $O_1 \ldots O_T$
2. Guess your first λ estimate $\lambda(0)$, k=0
3. k = k+1
4. Given $O_1 \ldots O_T$, $\lambda(k)$ compute
   $\gamma_t(i)$, $\varepsilon_t(i,j)$ $\quad \forall 1 \le t \le T, \quad \forall 1 \le i \le N, \quad \forall 1 \le j \le N$
5. Compute expected freq. of state i, and expected freq. $i \rightarrow j$
6. Compute new estimates of $\lambda(k+1)$: a_{ij}, $b_j(k)$, π_i accordingly.
7. Goto 3, unless converged.

EM for HMMs (BAUM-WELCH algorithm)

- Summary of BW-Algorithm
  - There are lots of local minima, which are usually adequate models of the data.
  - EM does not estimate the number of states.
    - Trade-off between too few states (inadequately modeling the structure in the data) and too many (fitting the noise).
    - Thus #states is a regularization parameter.
  - Often, HMMs are forced to have some links with zero probability.
    - This is done by setting a_{ij}=0 in initial estimate $\lambda(0)$

Seminar S05

- **Topics**
  - Gambler, or Speech Recognition by HMM
  - Viterbi Algorithm Exercise (see next time)
  - Viterbi Algorithm Application
  - Analyze HMM ***Forward*** or ***Backward*** estimation ***Recursively***. Illustrate them by completing the example in textbook
- **Requirements**
  - Prepare a **15 minutes** talk on your chosen topic
  - Make **ppt** to assist your talk
  - Prepare **at least 3 questions** to ask the audience after your talk
  - Get ready to **take questions** from the audience

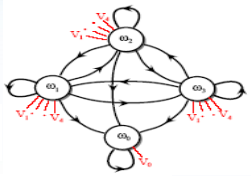
Exercise: Find the MPP

Observations: $V^4 = \{v_1, v_3, v_2, v_0\}$

Initial status: w1

$\pi_0 = a_{10} = 0.2, \quad \pi_1 = a_{11} = 0.3,$

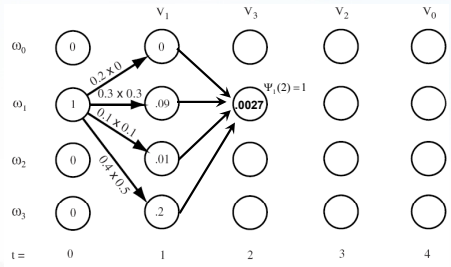
$\pi_2 = a_{12} = 0.1, \quad \pi_3 = a_{13} = 0.4$



$$a_{ij} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.2 & 0.3 & 0.1 & 0.4 \\ 0.2 & 0.5 & 0.2 & 0.1 \\ 0.8 & 0.1 & 0.0 & 0.1 \end{pmatrix} \qquad b_{jk} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0.4 & 0.1 & 0.2 \\ 0 & 0.1 & 0.1 & 0.7 & 0.1 \\ 0 & 0.5 & 0.2 & 0.1 & 0.2 \end{pmatrix}$$

Exercise: Find the MPP

- Complete this Trellis



Questions?