Biostatistics 615/815 Lecture 10: Hidden Markov Models

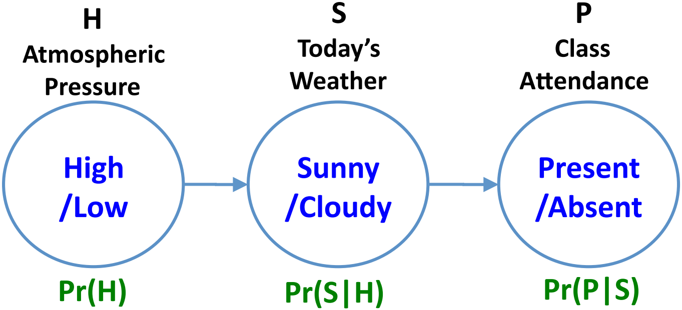
Invited Lecturer : Goo Jun

October 11th, 2011

Graphical Model 101

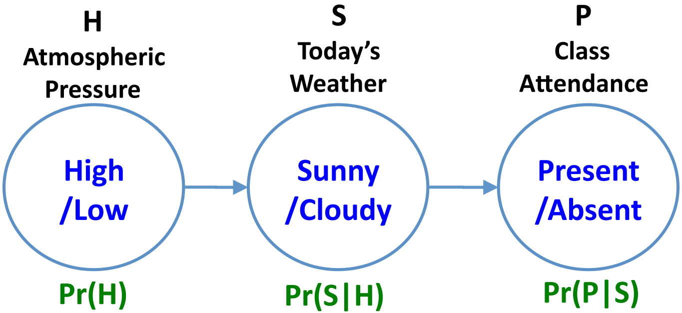
- Graphical model is marriage between probability theory and graph theory (Michiael I. Jordan)
- Each random variable is represented as vertex
- Dependency between random variables is modeled as edge
  - Directed edge : conditional distribution
  - Undirected edge : joint distribution
- Unconnected pair of vertices (without path from one to another) is independent
- An effective tool to represent complex structure of dependence / independence between random variables.

An example graphical model



- Are H and P independent?

An example graphical model



- Are H and P independent?
- Are H and P independent given S?

Example probability distribution

$\Pr(H)$

Value (H)	Description (H)	$\Pr(H)$
0	Low	0.3
1	High	0.7

$\Pr(S|H)$

S	Description (S)	H	Description (H)	$\Pr(S|H)$
0	Cloudy	0	Low	0.7
1	Sunny	0	Low	0.3
0	Cloudy	1	High	0.1
1	Sunny	1	High	0.9

Probability distribution (cont'd)

$\Pr(P|S)$

P	Description (P)	S	Description (S)	$\Pr(P\|S)$
0	Absent	0	Cloudy	0.5
1	Present	0	Cloudy	0.5
0	Absent	1	Sunny	0.1
1	Present	1	Sunny	0.9

Full joint distribution

$\Pr(H, S, P)$

H	S	P	$\Pr(H, S, P)$
0	0	0	0.105
0	0	1	0.105
0	1	0	0.009
0	1	1	0.081
1	0	0	0.035
1	0	1	0.035
1	1	0	0.063
1	1	1	0.567

- With a full join distribution, any type of inference is possible
- As the number of variables grows, the size of full distribution table increases exponentially

$\Pr(H, P|S) = \Pr(H|S)\Pr(P|S)$

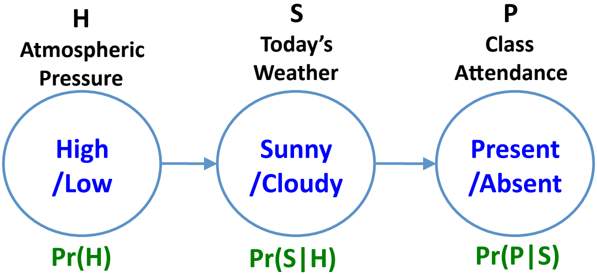
$\Pr(H, P|S)$

H	P	S	$\Pr(H, P\|S)$
0	0	0	0.3750
0	1	0	0.3750
1	0	0	0.1250
1	1	0	0.1250
0	0	1	0.0125
0	1	1	0.1125
1	0	1	0.0875
1	1	1	0.7875

$\Pr(H|S)$, $\Pr(P|S)$

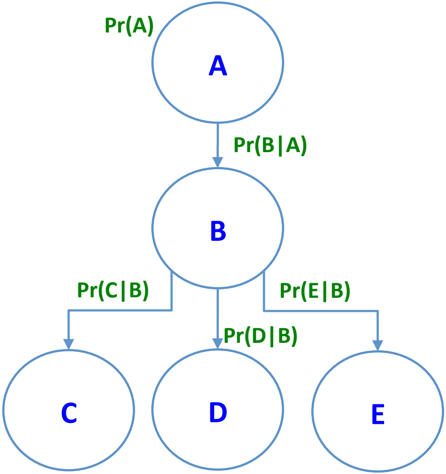
H	S	$\Pr(H\|S)$	P	S	$\Pr(P\|S)$
0	0	0.750	0	0	0.500
1	0	0.250	1	0	0.500
0	1	0.125	0	1	0.100
1	1	0.875	1	1	0.900

H and P are conditionally independent given S



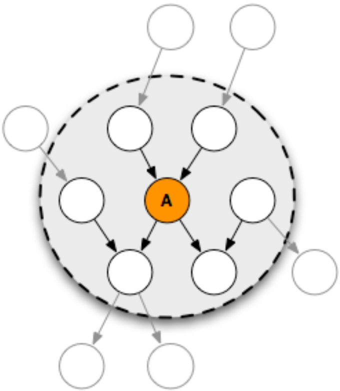
- H and P do not have direct path one from another
- All path from H to P is connected thru S.
- Conditioning on S separates H and P

Conditional independence in graphical models



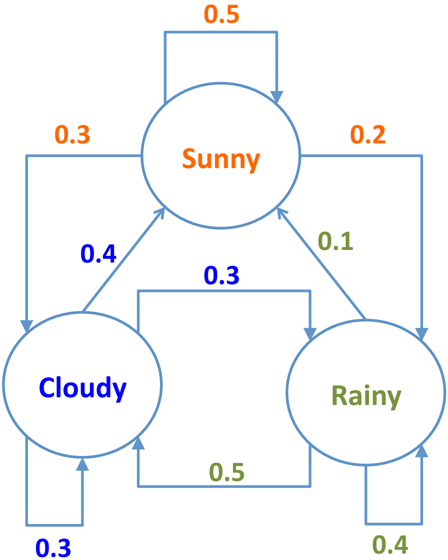
- $\Pr(A, C, D, E|B) = \Pr(A|B)\Pr(C|B)\Pr(D|B)\Pr(E|B)$

Markov Blanket



- If conditioned on the variables in the gray area (variables with direct dependency), A is independent of all the other nodes.
- $A \perp (U - A - \pi_A) | \pi_A$

Markov Process : An example



Mathematical representation of a Markov Process

$$\pi = \begin{pmatrix} \Pr(q_1 = S_1 = \text{Sunny}) \\ \Pr(q_1 = S_2 = \text{Cloudy}) \\ \Pr(q_1 = S_3 = \text{Rainy}) \end{pmatrix} = \begin{pmatrix} 0.7 \\ 0.2 \\ 0.1 \end{pmatrix}$$

$$A_{ij} = \Pr(q_{t+1} = S_j | q_t = S_i)$$

$$A = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.4 & 0.3 & 0.3 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}$$

Example questions in Markov Process

What is the chance of rain in the day 2?

Example questions in Markov Process

What is the chance of rain in the day 2?

$$\Pr(q_2 = S_3) = (A^T \pi)_3 = 0.24$$

Example questions in Markov Process

What is the chance of rain in the day 2?

$$\Pr(q_2 = S_3) = (A^T \pi)_3 = 0.24$$

If it rains today, what is the chance of rain on the day after tomorrow?

Example questions in Markov Process

What is the chance of rain in the day 2?

$$\Pr(q_2 = S_3) = (A^T \pi)_3 = 0.24$$

If it rains today, what is the chance of rain on the day after tomorrow?

$$\Pr(q_3 = S_3 | q_1 = S_3) = \left[(A^T)^2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right]_3 = 0.33$$

Example questions in Markov Process

What is the chance of rain in the day 2?

$$\Pr(q_2 = S_3) = (A^T \pi)_3 = 0.24$$

If it rains today, what is the chance of rain on the day after tomorrow?

$$\Pr(q_3 = S_3 | q_1 = S_3) = \left[(A^T)^2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right]_3 = 0.33$$

Stationary distribution

$$\begin{aligned} \mathbf{p} &= A^T \mathbf{p} \\ p &= (0.346, 0.359, 0.295)^T \end{aligned}$$

Markov process is only dependent on the previous state

If it rains today, what is the chance of rain on the day after tomorrow?

$$\Pr(q_3 = S_3 | q_1 = S_3) = \left[(A^T)^2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right]_3 = 0.33$$

Markov process is only dependent on the previous state

If it rains today, what is the chance of rain on the day after tomorrow?

$$\Pr(q_3 = S_3 | q_1 = S_3) = \left[(A^T)^2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right]_3 = 0.33$$

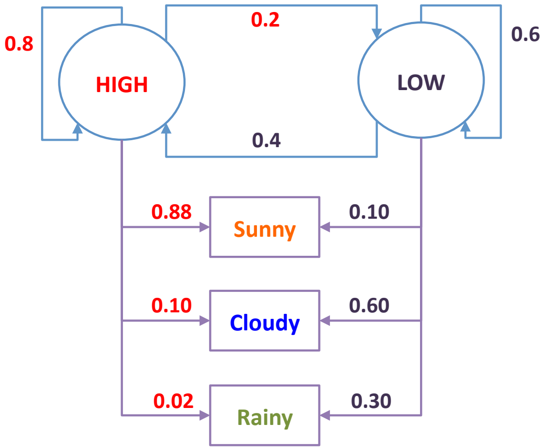
If it has rained for the past three days, what is the chance of rain on the day after tomorrow?

$$\Pr(q_5 = S_3 | q_1 = q_2 = q_3 = S_3) = \Pr(q_5 = S_3 | q_3 = S_3) = 0.33$$

Hidden Markov Models (HMMs)

- A Markov model where actual state is unobserved
  - Transition between states are probablistically modeled just like the Markov process
- Typically there are observable outputs associated with hidden states
  - The probability distribution of observable outputs given an hidden states can be obtained.

An example of HMM



- Direct Observation : (SUNNY, CLOUDY, RAINY)
- Hidden States : (HIGH, LOW)

Mathematical representation of the HMM example

States $S = \{S_1, S_2\} =$ (HIGH, LOW)

Mathematical representation of the HMM example

States $S = \{S_1, S_2\} = $ (HIGH, LOW)

Outcomes $O = \{O_1, O_2, O_3\} = $ (SUNNY, CLOUDY, RAINY)

Mathematical representation of the HMM example

States $S = \{S_1, S_2\} =$ (HIGH, LOW)

Outcomes $O = \{O_1, O_2, O_3\} =$ (SUNNY, CLOUDY, RAINY)

Initial States $\pi_i = \Pr(q_1 = S_i)$, $\pi = \{0.7, 0.3\}$

Mathematical representation of the HMM example

States $S = \{S_1, S_2\} =$ (HIGH, LOW)

Outcomes $O = \{O_1, O_2, O_3\} =$ (SUNNY, CLOUDY, RAINY)

Initial States $\pi_i = \Pr(q_1 = S_i)$, $\pi = \{0.7, 0.3\}$

Transition $A_{ij} = \Pr(q_{t+1} = S_j | q_t = S_i)$

$$A = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}$$

Emission $B_{ij} = b_{q_t}(o_t) = b_{S_i}(O_j) = \Pr(o_t = O_j | q_t = S_i)$

$$B = \begin{pmatrix} 0.88 & 0.10 & 0.02 \\ 0.10 & 0.60 & 0.30 \end{pmatrix}$$

More Markov Chain Questions

- Marginal probability : What is the chance of rain in the day 4?

More Markov Chain Questions

- Marginal probability : What is the chance of rain in the day 4?
- Conditioned to previous observations : What is the chance of rain in the day 2, if it rained in the day 1?

More Markov Chain Questions

- Marginal probability : What is the chance of rain in the day 4?
- Conditioned to previous observations : What is the chance of rain in the day 2, if it rained in the day 1?
- Forward-backward algorithms : If the observation was (SUNNY,SUNNY,CLOUDY,RAINY,RAINY) from day 1 through day 5, what is the distribution of hidden states for each day?

More Markov Chain Questions

- Marginal probability : What is the chance of rain in the day 4?
- Conditioned to previous observations : What is the chance of rain in the day 2, if it rained in the day 1?
- Forward-backward algorithms : If the observation was (SUNNY,SUNNY,CLOUDY,RAINY,RAINY) from day 1 through day 5, what is the distribution of hidden states for each day?
- Viterbi algorithm : If the observation was (SUNNY,SUNNY,CLOUDY,RAINY,RAINY) from day 1 through day 5, what would be the mostly likely sequence of states?

Unconditional marginal probabilities

What is the chance of rain in the day 4?

$$\mathbf{f}(\mathbf{q}_3) = \begin{pmatrix} \Pr(q_4 = S_1) \\ \Pr(q_4 = S_2) \end{pmatrix} = (A^T)^3\pi = \begin{pmatrix} 0.669 \\ 0.331 \end{pmatrix}$$

$$\mathbf{g}(o_4) = \begin{pmatrix} \Pr(o_4 = O_1) \\ \Pr(o_4 = O_2) \\ \Pr(o_4 = O_3) \end{pmatrix} = B^T\mathbf{f}(\mathbf{q}_4) = \begin{pmatrix} 0.621 \\ 0.266 \\ 0.233 \end{pmatrix}$$

The chance of rain in day 3 is 23.3%

Marginal likelihood of data in HMM

- Let $\lambda = (A, B, \pi)$
- For a sequence of observation $\mathbf{o} = \{o_1, \cdots, o_t\}$,

$$\Pr(\mathbf{o}|\lambda) = \sum_{\mathbf{q}} \Pr(\mathbf{o}|\mathbf{q}, \lambda) \Pr(\mathbf{q}|\lambda)$$

$$\Pr(\mathbf{o}|\mathbf{q}, \lambda) = \prod_{i=1}^{t} \Pr(o_i|q_i, \lambda) = \prod_{i=1}^{t} b_{q_i}(o_i)$$

$$\Pr(\mathbf{q}|\lambda) = \pi_{q_1} \prod_{i=2}^{t} a_{q_{i-1} q_i}$$

$$\Pr(\mathbf{o}|\lambda) = \sum_{\mathbf{q}} \pi_{q_1} b_{q_1}(o_1) \prod_{i=2}^{t} a_{q_{i-1} q_i} b_{q_i}(o_i)$$

Naive computation of the likelihood

$$\Pr(\mathbf{o}|\lambda) = \sum_{\mathbf{q}} \pi_{q_1} b_{q_1}(o_1) \prod_{i=2}^{t} a_{q_{i-1}q_i} b_{q_i}(o_i)$$

- Number of possible $q = 2^t$ are exponentially growing with the number of observations
- Computational would be infeasible for large number of observations
- Algorithmic solution required for efficient computation.

More Markov Chain Question

- If the observation was (SUNNY,SUNNY,CLOUDY,RAINY,RAINY) from day 1 through day 5, what is the distribution of hidden states for each day?
- Need to know $\Pr(q_t|\mathbf{o}, \lambda)$

Forward and backward probabilities

$$\begin{aligned}
\mathbf{q}_t^- &= (q_1, \cdots, q_{t-1}), \qquad \mathbf{q}_t^+ = (q_{t+1}, \cdots, q_T) \\
\mathbf{o}_t^- &= (o_1, \cdots, o_{t-1}), \qquad \mathbf{o}_t^+ = (o_{t+1}, \cdots, o_T) \\
\Pr(q_t = i|\mathbf{o}, \lambda) &= \frac{\Pr(q_t = i, \mathbf{o}|\lambda)}{\Pr(\mathbf{o}|\lambda)} = \frac{\Pr(q_t = i, \mathbf{o}|\lambda)}{\sum_{j=1}^{n} \Pr(q_t = j, \mathbf{o}|\lambda)} \\
\Pr(q_t, \mathbf{o}|\lambda) &= \Pr(q_t, \mathbf{o}_t^-, o_t, \mathbf{o}_t^+|\lambda) \\
&= \Pr(\mathbf{o}_t^+|q_t, \lambda) \Pr(\mathbf{o}_t^-|q_t, \lambda) \Pr(o_t|q_t, \lambda) \Pr(q_t|\lambda) \\
&= \Pr(\mathbf{o}_t^+|q_t, \lambda) \Pr(\mathbf{o}_t^-, o_t, q_t|\lambda) \\
&= \beta_t(q_t)\alpha_t(q_t)
\end{aligned}$$

If $\alpha_t(q_t)$ and $\beta_t(q_t)$ is known, $\Pr(q_t|\mathbf{o}, \lambda)$ can be computed in a linear time.

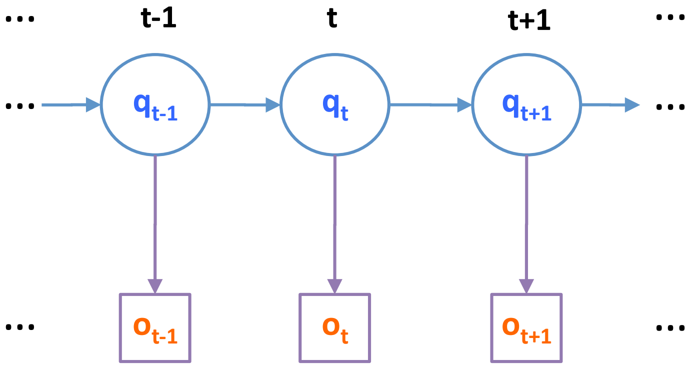
DP algorithm for calculating forward probability

- Key idea is to use $(q_t, o_t) \perp \mathbf{o}_t^- | \mathbf{q}_{t-1}$.
- Each of q_{t-1}, q_t, and q_{t+1} is a Markov blanket.

$$
\begin{aligned}
\alpha_t(i) &= \Pr(o_1, \cdots, o_t, q_t = i|\lambda) \\
&= \sum_{j=1}^{n} \Pr(\mathbf{o}_t^-, o_t, q_{t-1} = j, q_t = i|\lambda) \\
&= \sum_{j=1}^{n} \Pr(\mathbf{o}_t^-, q_{t-1} = j|\lambda) \Pr(q_t = i|q_{t-1} = j, \lambda) \Pr(o_t|q_t = i, \lambda) \\
&= \sum_{j=1}^{n} \alpha_{t-1}(j) a_{ji} b_i(o_t) \\
\alpha_1(i) &= \pi_i b_i(o_1)
\end{aligned}
$$

Conditional dependency in forward-backward algorithms

- Forward : $(q_t, o_t) \perp \mathbf{o}_t^- | \mathbf{q}_{t-1}$.
- Backward : $o_{t+1} \perp \mathbf{o}_{t+1}^+ | \mathbf{q}_{t+1}$.



DP algorithm for calculating backward probability

- Key idea is to use $o_{t+1} \perp \mathbf{o}_{t+1}^{+} | \mathbf{q}_{t+1}$.

$$\begin{aligned}
\beta_t(i) &= \Pr(o_{t+1}, \cdots, o_T | q_t = i, \lambda) \\
&= \sum_{j=1}^{n} \Pr(o_{t+1}, \mathbf{o}_{t+1}^{+}, q_{t+1} = j | q_t = i, \lambda) \\
&= \sum_{j=1}^{n} \Pr(o_{t+1} | q_{t+1}, \lambda) \Pr(\mathbf{o}_{t+1}^{+} | q_{t+1} = j, \lambda) \Pr(q_{t+1} = j | q_t = i, \lambda) \\
&= \sum_{j=1}^{n} \beta_{t+1}(j) a_{ij} b_j(o_{t+1}) \\
\beta_T(i) &= 1
\end{aligned}$$

Putting forward and backward probabilities together

- Conditional probability of states given data

$$\Pr(q_t = i|\mathbf{o}, \lambda) = \frac{\Pr(\mathbf{o}, q_t = S_i|\lambda)}{\sum_{j=1}^{n} \Pr(\mathbf{o}, q_t = S_j|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{n} \alpha_t(j)\beta_t(j)}$$

- Time complexity is $\Theta(n^2 T)$.

Finding the most likely trajectory of hidden states

- Given a series of observations, we want to compute

$$\arg\max_{\mathbf{q}} \Pr(\mathbf{q}|\mathbf{o}, \lambda)$$

- Define $\delta_t(i)$ as

$$\delta_t(i) = \max_{\mathbf{q}} \Pr(\mathbf{q}, \mathbf{o}|\lambda)$$

- Use dynamic programming algorithm to find the 'most likely' path

The Viterbi algorithm

Initialization $\delta_1(i) = \pi b_i(o_1)$ for $1 \leq i \leq n$.

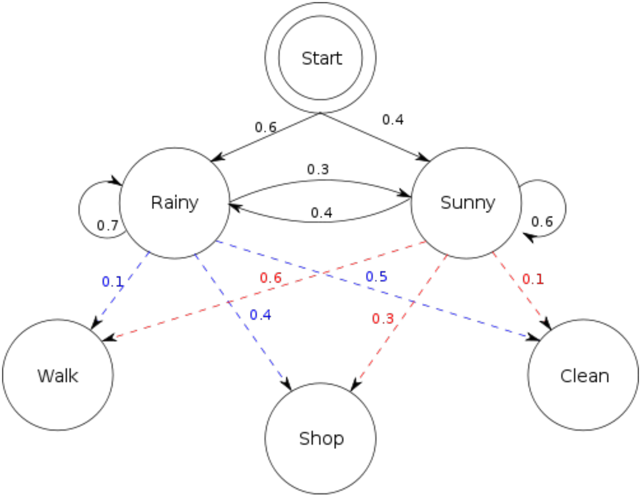
Maintenance $\delta_t(i) = \max_j \delta_{t-1}(j) a_{ji} b_i(o_t)$

$\phi_t(i) = \arg\max_j \delta_{t-1}(j) a_{ji}$

Termination Max likelihood is $\max_i \delta_T(i)$

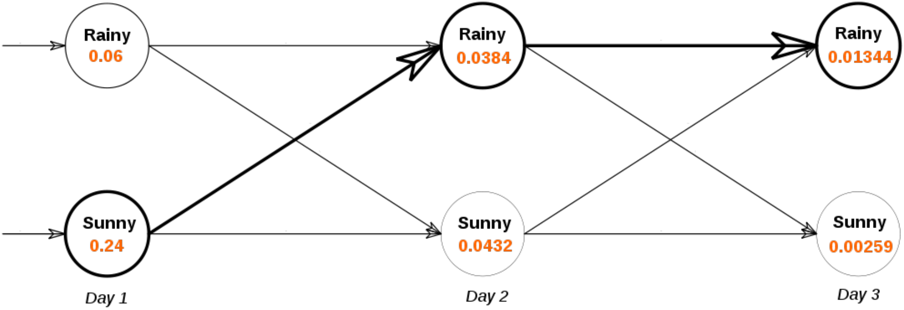
Optimal path can be backtracked using $\phi_t(i)$

An HMM example



An example Viterbi path

- When observations were (walk, shop, clean)
- Similar to Manhattan tourist problem.



Summary

Today - Hidden Markov Models

- Graphical models and conditional independence
- Forward-backward algorithm
- Viterbi algorithm

Next lectures

- Implementations of hidden Markov Models