HMKANG@UMICH.EDU

MODULE 1 / UNIT 9

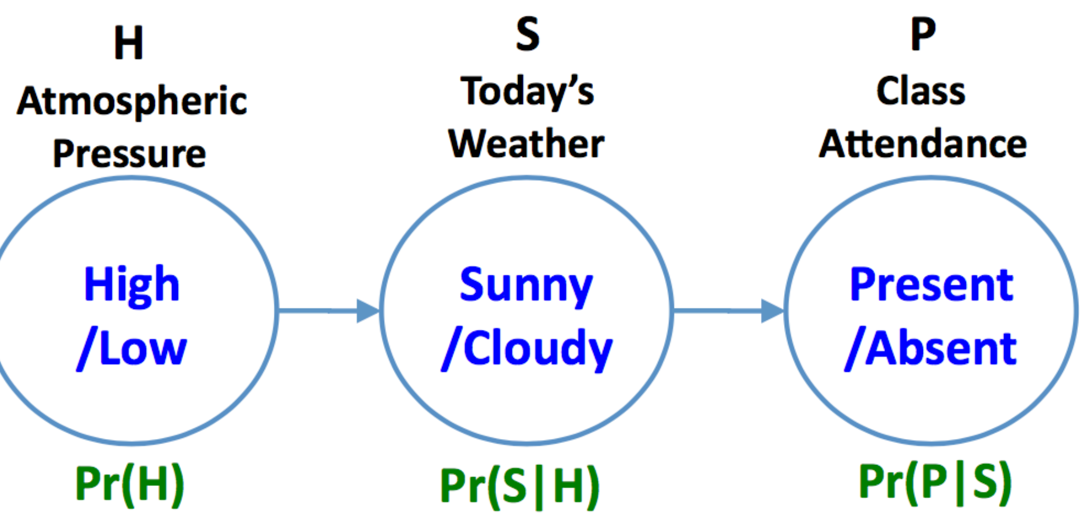
HIDDEN MARKOV MODELS

Outline

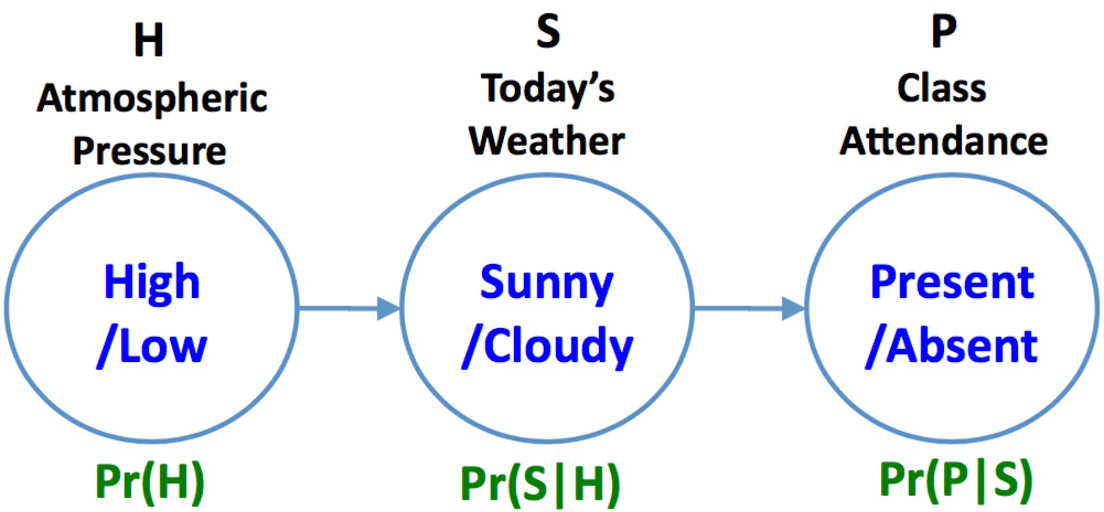
- Graphical models
- Markov process
- Hidden Markov Models (HMM)
- Forward-backward algorithm
- Viterbi algorithm

Graphical Models 101

- Marriage between probability theory and graph theory
- An effective tool to represent complex structure of dependence/independence between random variable
- Vertex : a random variable
- Edge : dependency between random variables



An example graphical model



Is $H \perp P$? (marginal independence)

Is $H \perp P \mid S$? (conditional independence)

An example probability distribution

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

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

	P		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

The full joint distribution

H		$Pr(H)$
0	Low	0.3
1	High	0.7

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

S		H		$Pr(S/H)$
0	Cloudy	0	Low	0.7
1	High	0	Low	0.3
0	Cloudy	1	High	0.1
1	High	1	High	0.9

P		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

Conditional distribution

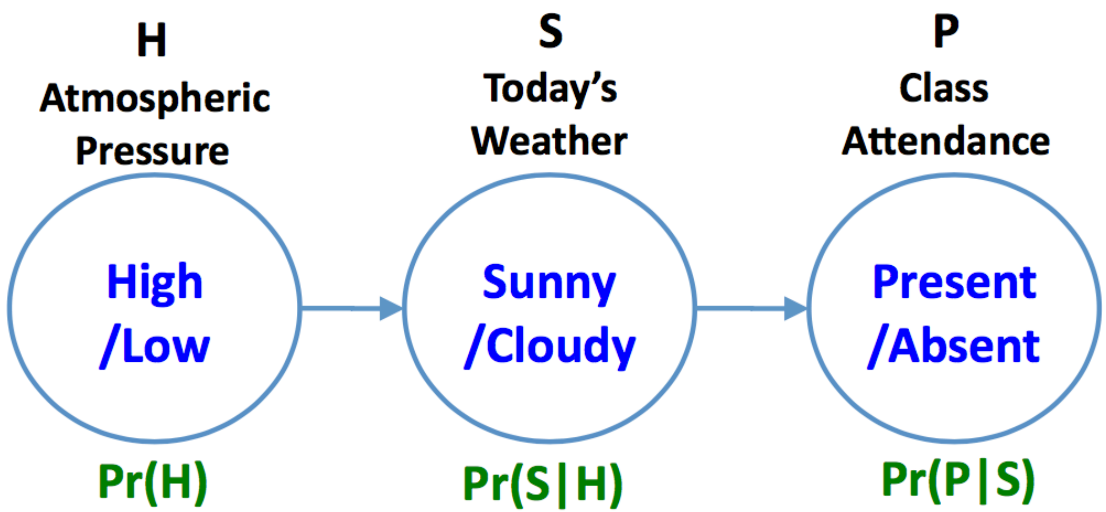
Is $H \perp P \mid S$?

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

H		S		Pr(H/S)
0	Low	0	Cloudy	0.750
1	High	0	Cloudy	0.250
0	Low	1	Sunny	0.125
1	High	1	Sunny	0.875

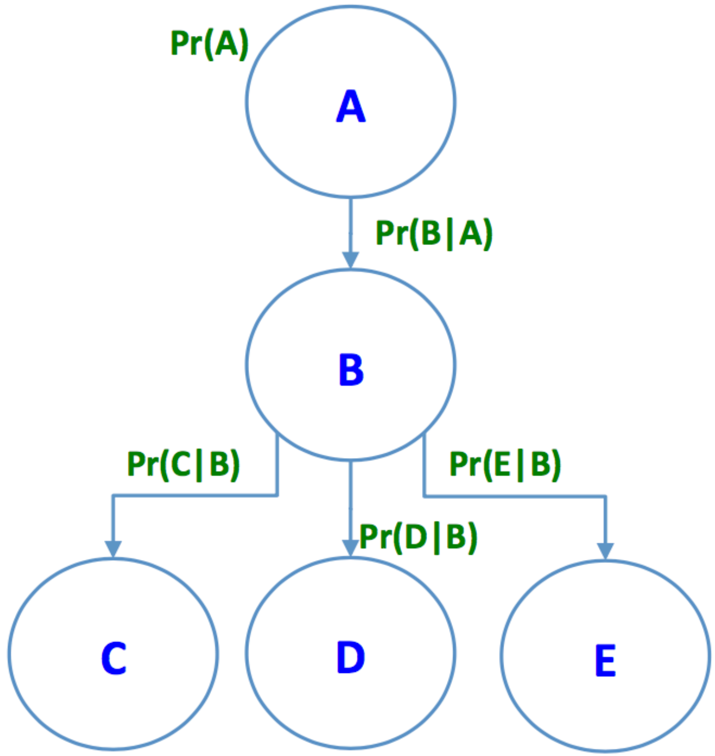
P		S		Pr(P/S)
0	Absent	0	Cloudy	0.5
1	Present	0	Cloudy	0.5
0	Absent	1	Sunny	0.1
1	Present	1	Sunny	0.9

H and *P* are conditionally independent given *S*



- *H* and *P* does not have direct path in the graph
- All paths from *H* to *P* are connected through *S*
- In such cases, conditioning on *S* separates *H* and *P*

More conditional independence



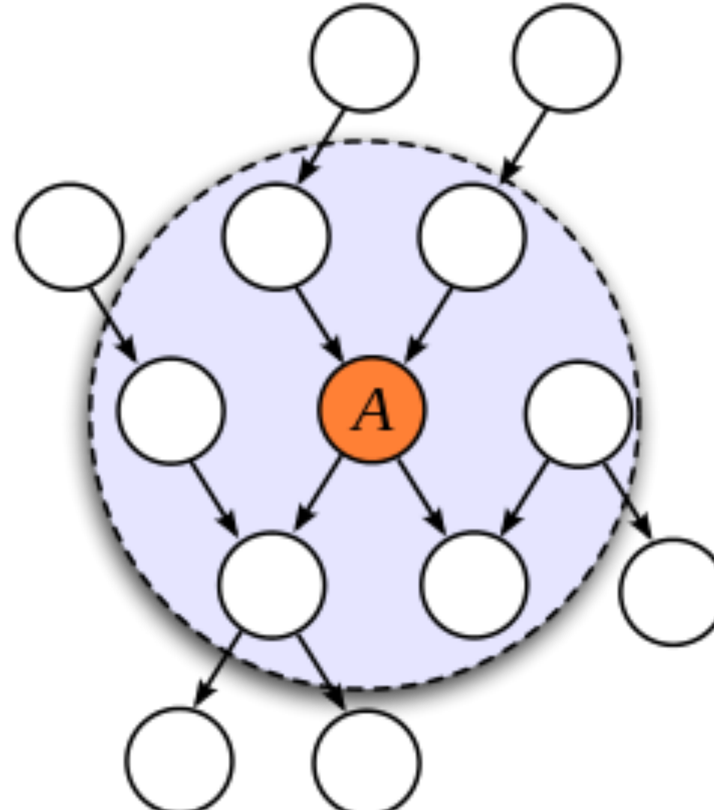
$$\Pr(A, C, D, E | B) = \Pr(A|B) \Pr(C|B) \Pr(D|B) \Pr(E|B)$$

Markov blanket

- Markov blanket ∂A for node A is a set of nodes composed of A's parents, children, and the other parents of its children.

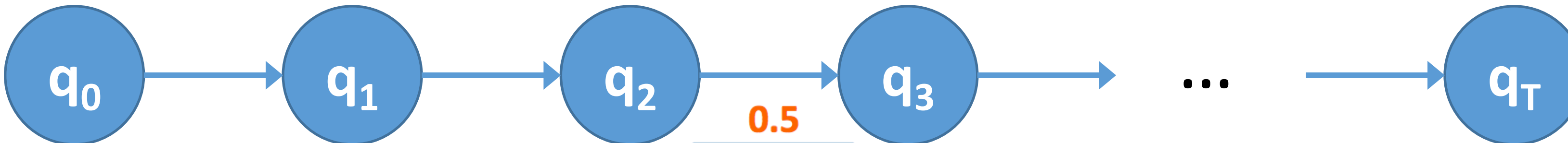
- Fact : Node A is conditionally independent given its Markov blanket ∂A.

$$A \perp (U - A - \partial A) \mid \partial A$$



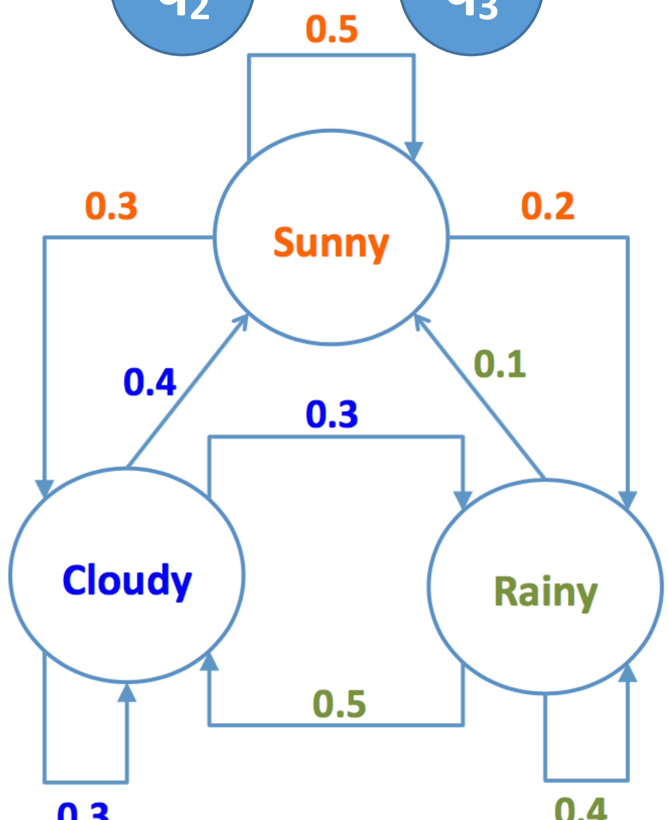
https://en.wikipedia.org/wiki/Markov_blanket

Markov process



State diagram

$\Pr(q_{t+1}|q_t)$



Markov process : mathematical representation

$$
\begin{aligned}
\pi &= \begin{pmatrix} \Pr(q_1 = S_1 = \text{Sunny}) \\ \Pr(q_1 = S_2 = \text{Cloudy}) \\ \Pr(q_1 = S_3 = \text{Rainy}) \end{pmatrix} = \begin{pmatrix} 0.7 \\ 0.2 \\ 0.1 \end{pmatrix} \\
A_{ij} &= \Pr(q_{t+1} = S_j | q_t = S_i) \\
A &= \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.4 & 0.3 & 0.3 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}
\end{aligned}
$$

Markov process : example questions

- What is the chance of rain in day 2?

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

- What is the stationary distribution when T

Markov process : example questions

- **What is the chance of rain in 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?**

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

- **What is the stationary distribution when t is very large?**

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

Markov process dependencies

- 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 3 days in a row, what is the chance of rain on the day after tomorrow?

Markov process dependencies

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

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

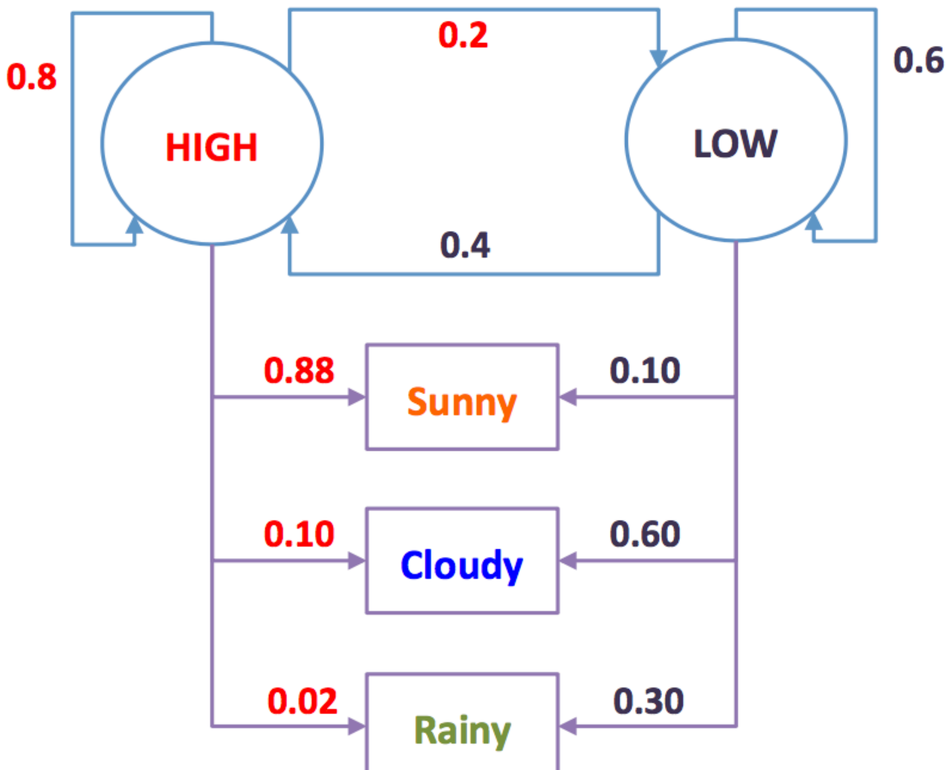
- If it has rained for the past 3 days in a row, what is the chance of rain on the day after tomorrow?

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

Hidden Markov Models (HMMs)

- A graphical model composed of hidden states and observed data.
- The hidden states follow a Markov process
- The distribution of observed data depends only on the hidden state.

An example of HMM



- **Hidden states:**
  - HIGH
  - LOW
- **Observed data**
  - Sunny
  - Cloudy
  - Rainy

Mathematical representation of the example

- States $S = \{S_1, S_2\} = (\text{HIGH, LOW})$
- Observed Data $O = \{O_1, O_2, O_3\} = (\text{SUNNY, CLOUDY, RAINY})$
- 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 $B_{ij} = b_{q_t}(o_t) = b_{S_i}(O_j) = \Pr(o_t = O_j | q_t = S_i)$

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

Marginal probabilities

- **What is the chance of rain in the day 4?**

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

$$\begin{pmatrix} \Pr(o_4 = O_1) \\ \Pr(o_4 = O_2) \\ \Pr(o_4 = O_3) \end{pmatrix} = B^T (A^T)^3 \pi = \begin{pmatrix} 0.621 \\ 0.266 \\ 0.233 \end{pmatrix}$$

Conditional probabilities

- **If the observation was** (SUNNY, SUNNY, CLOUDY, RAINY, RAINY) = $(O_1, O_1, O_2, O_3, O_3)$ **from day 1 to day 5, what is the distribution of the hidden states on each day?**

$$\Pr\left(q_i | o_1, o_2, o_3, o_4, o_5; \lambda = (\pi, A, B)\right)$$

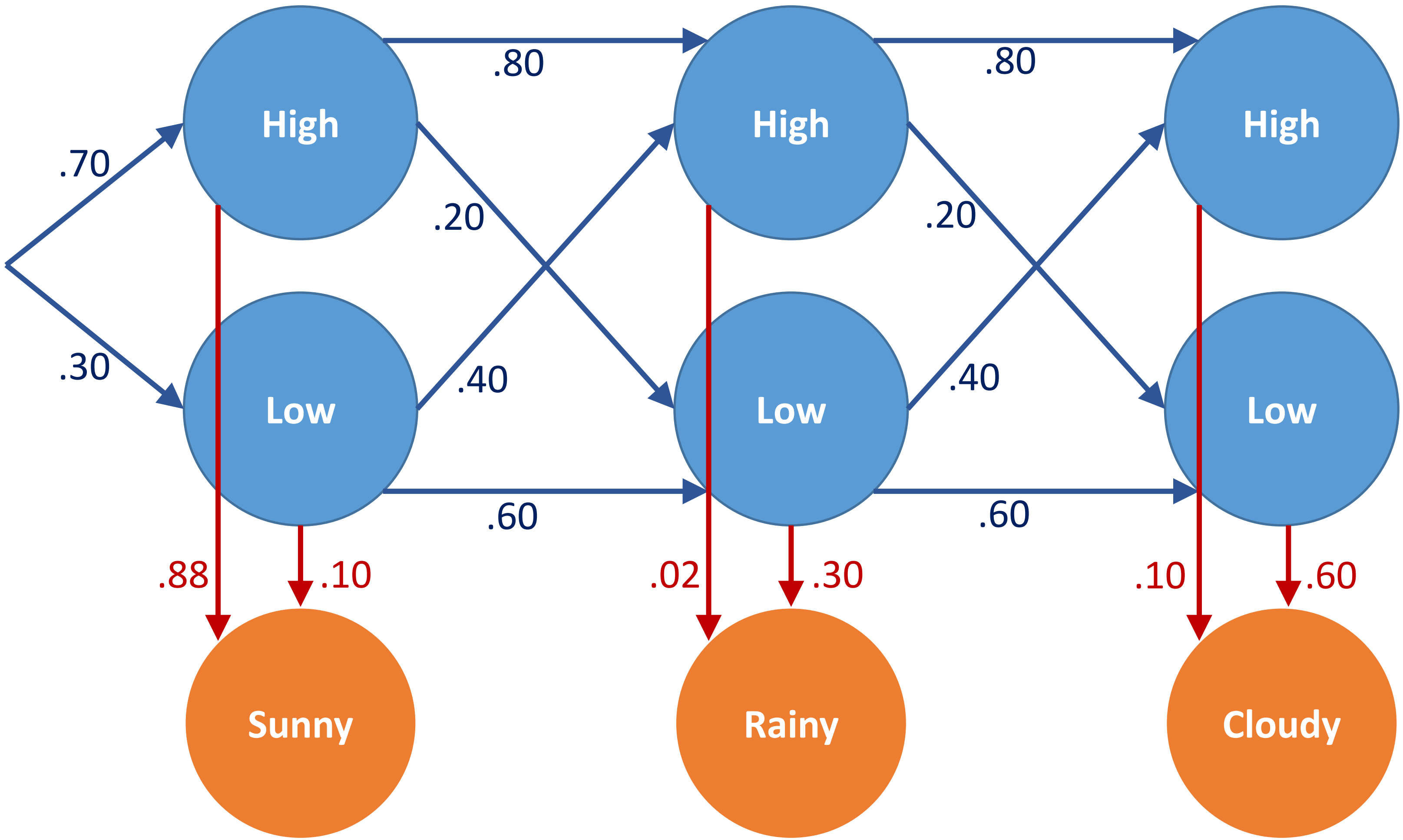
$$(1 \leq i \leq 5)$$

Most likely states

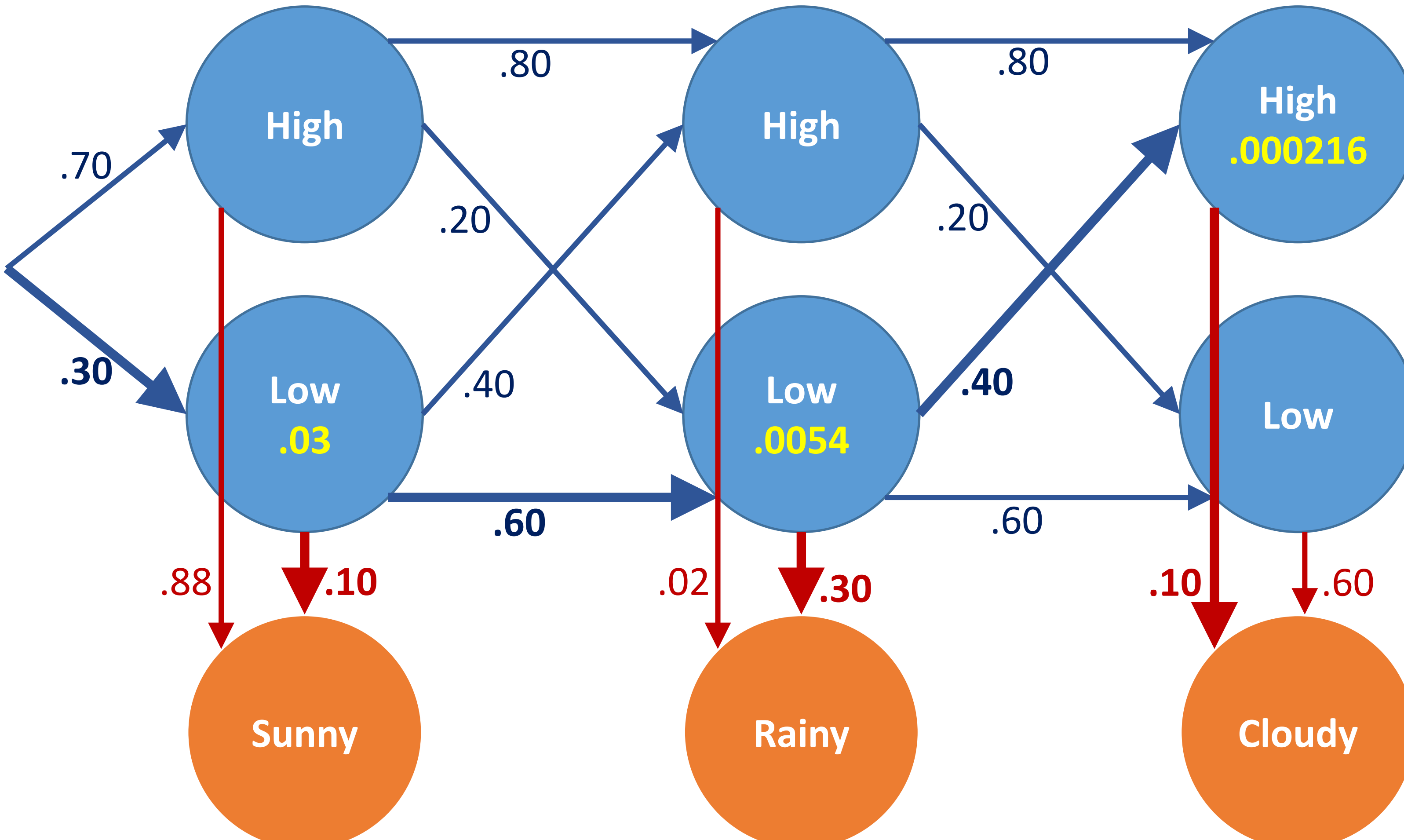
- **If the observation was** (SUNNY, SUNNY, CLOUDY, RAINY, RAINY) = $(O_1, O_1, O_2, O_3, O_3)$ **from day 1 to day 5, what is the most likely combinations of the hidden states?**

$$\arg\max_{(q_1,\ldots,q_5)} \Pr\left(q_1, \ldots, q_5 \middle| o_1, \ldots, o_5; \lambda = (\pi, A, B)\right)$$

Finding MLE states



Enumerating a possible state



Enumerating all possible states

q_{111}	q_2	q_3	Pr(q)	$Pr(o_1 \mid q_1)$	$Pr(o_2 \mid q_2)$	$Pr(o_3 \mid q_3)$	Pr(q,o)
H	H	H	.448	.88	.02	.10	.0008
H	H	L	.112	.88	.02	.60	.0012
H	L	H	.056	.88	.30	.10	.0015
H	**L**	**L**	**.084**	**.88**	**.30**	**.60**	**.0133**
L	H	H	.096	.10	.02	.10	<.0001
L	H	L	.024	.10	.02	.60	<.0001
L	L	H	.072	.10	.30	.10	.0002
L	L	L	.108	.10	.30	.60	.0019

Brute-force calculation of MLE states

- Enumerating all possible combination of states takes $O(2^T)$.
- Is this reasonable fast or too slow? Why?
- Is there a way to reduce the time complexity? How?