Lecture 1:

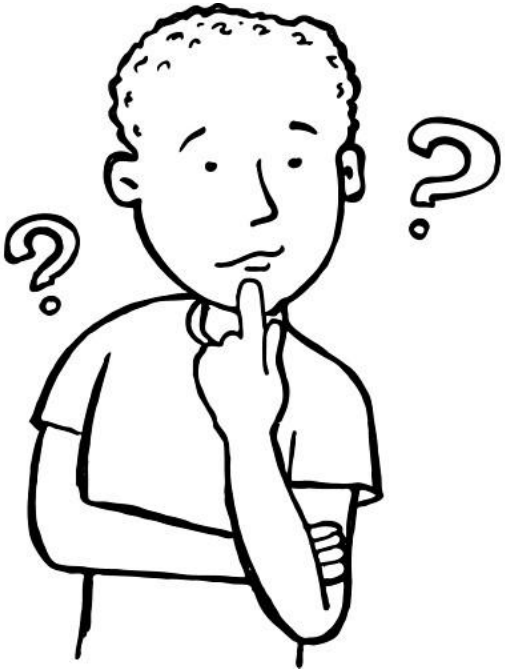
Introduction to Markov Models

Outline

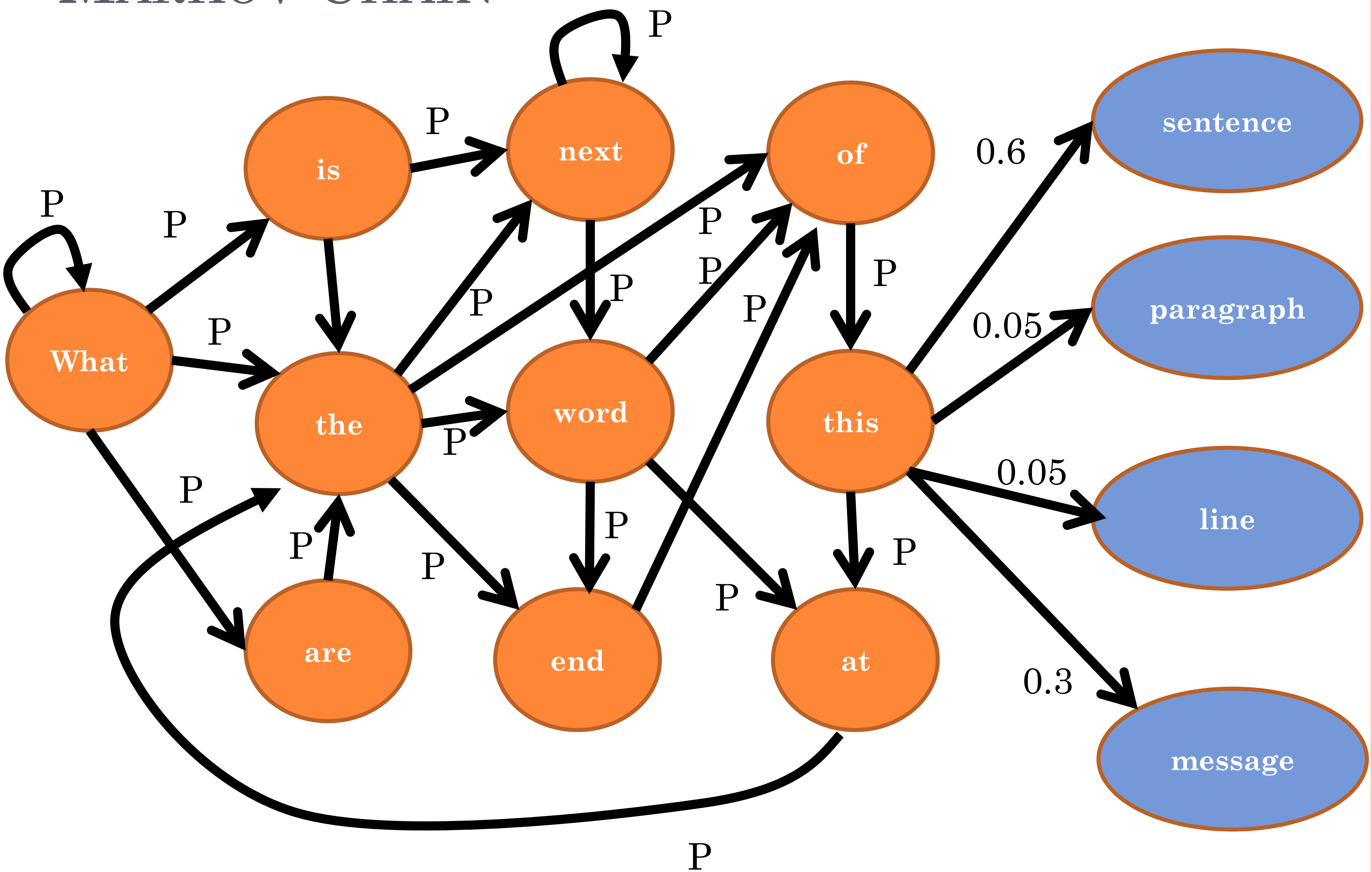
- Markov model
- Hidden Markov model (HMM)
- Example: dice & coins
- Example: recognizing eating activities

Motivation

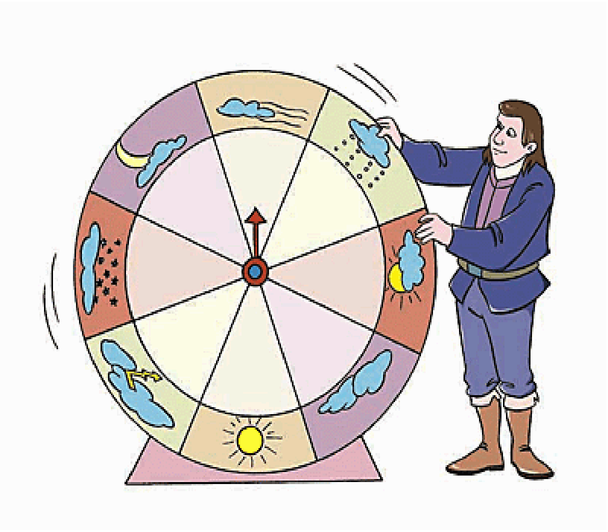
What is the word
at the end of this _____?



Markov Chain

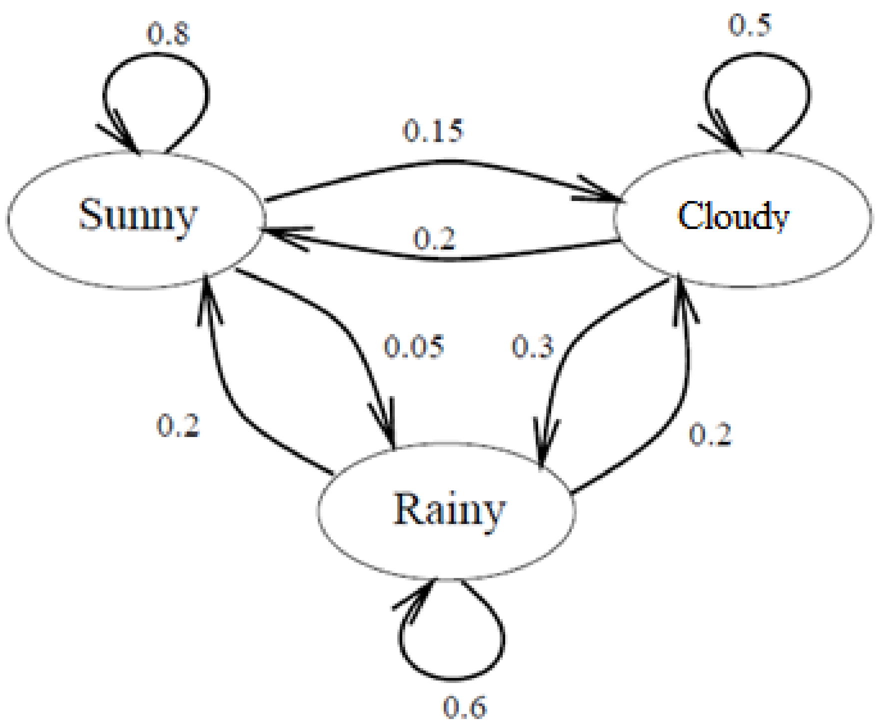


Markov Chain: Weather Example



- Design a Markov Chain to predict the weather of tomorrow using previous information of the past days.
- Our model has only 3 states: $S = \{S_1, S_2, S_3\}$, and the name of each state is $S_1 = Sunny$, $S_2 = Rainy, S_3 = Cloudy$.
- To establish the transition probabilities relationship between states we will need to collect data.

- Assume the data produces the following transition probabilities:



$$\left.\begin{array}{r} P(Sunny|Sunny) = 0.8 \\ P(Rainy|Sunny) = 0.05 \\ P(Cloudy|Sunny) = 0.15 \end{array}\right\} 1$$

$$\left.\begin{array}{r} P(Sunny|Rainy) = 0.2 \\ P(Rainy|Rainy) = 0.6 \\ P(Cloudyy|Rainy) = 0.2 \end{array}\right\} 1$$

$$\left.\begin{array}{r} P(Sunny|Cloudy) = 0.2 \\ P(Rainy|Cloudy) = 0.3 \\ P(Cloudy|Cloudy) = 0.5 \end{array}\right\} 1$$

- Let's say we have a sequence: Sunny, Rainy, Cloudy, Cloudy, Sunny, Sunny, Sunny, Rainy, ….; so, in a day we can be in any of the three states.

- We can use the following **state sequence** notation: q_1, q_2, q_3, q_4, q_5, … . ., where $q_i \epsilon \{Sunny, Rainy, Cloudy\}$.

- In order to compute the probability of tomorrow's weather we can use the Markov property:

$$P(q_1, \dots, q_n) = \prod_{i=1}^{n} P(q_i | q_{i-1})$$

- Exercise 1: Given that today is Sunny, what's the probability that tomorrow is Sunny and the next day Rainy?

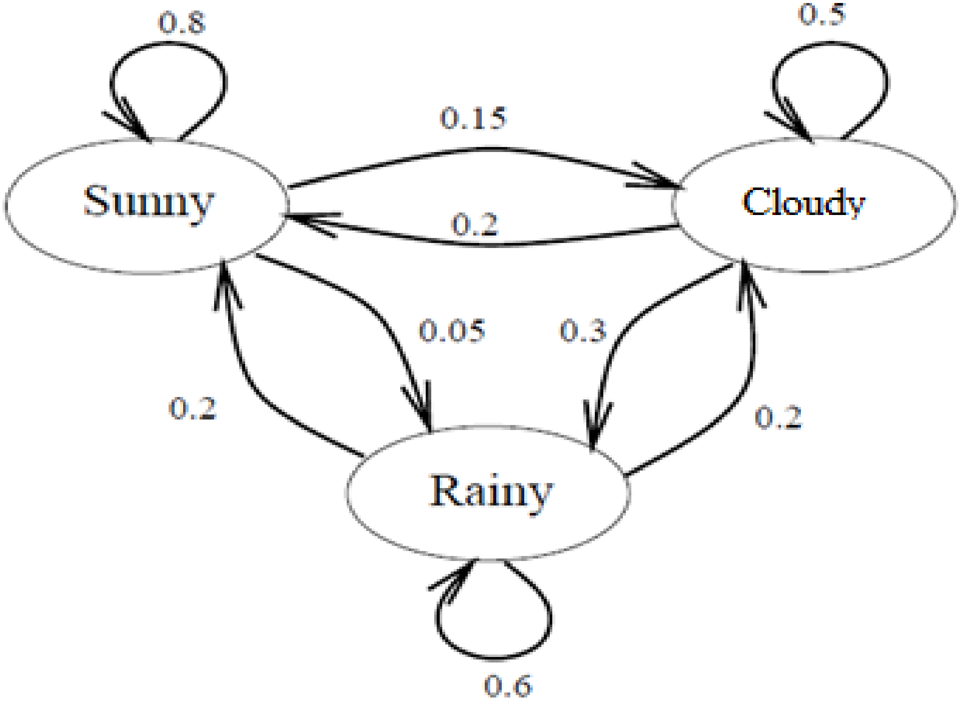
$P(q_2, q_3|q_1) = P(q_2|q_1)P(q_3|q_1, q_2)$

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

$= P(Sunny|Sunny)\ P(Rainy|Sunny)$

$= (0.8)(0.05)$

$= 0.04$

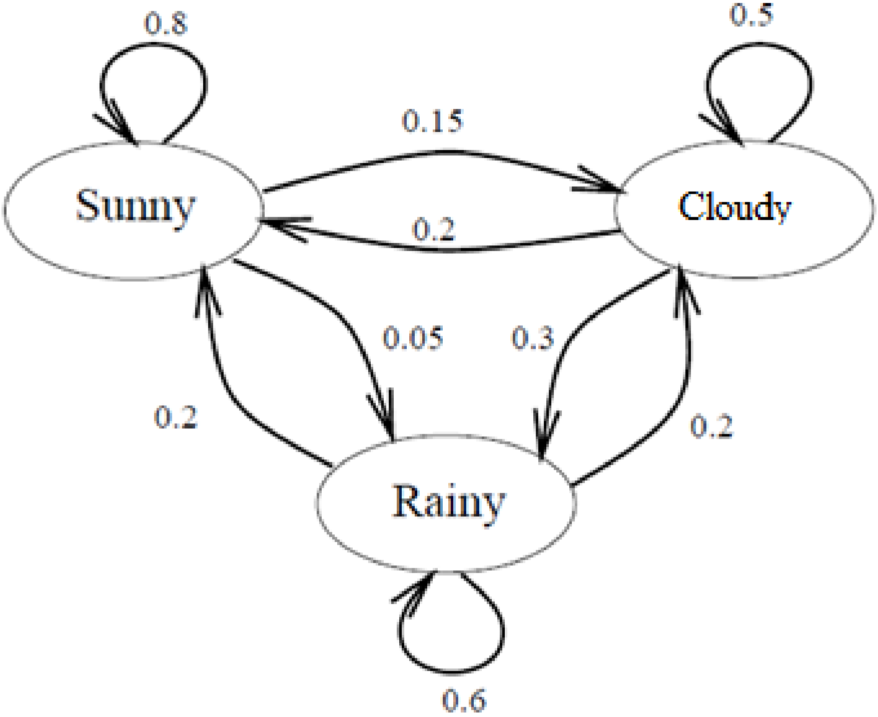


- Exercise 2: Assume that yesterday's weather was Rainy, and today is Cloudy, what is the probability that tomorrow will be Sunny?

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

$$= P(Sunny|Cloudy)$$

$$= 0.2$$



What is a Markov Model?

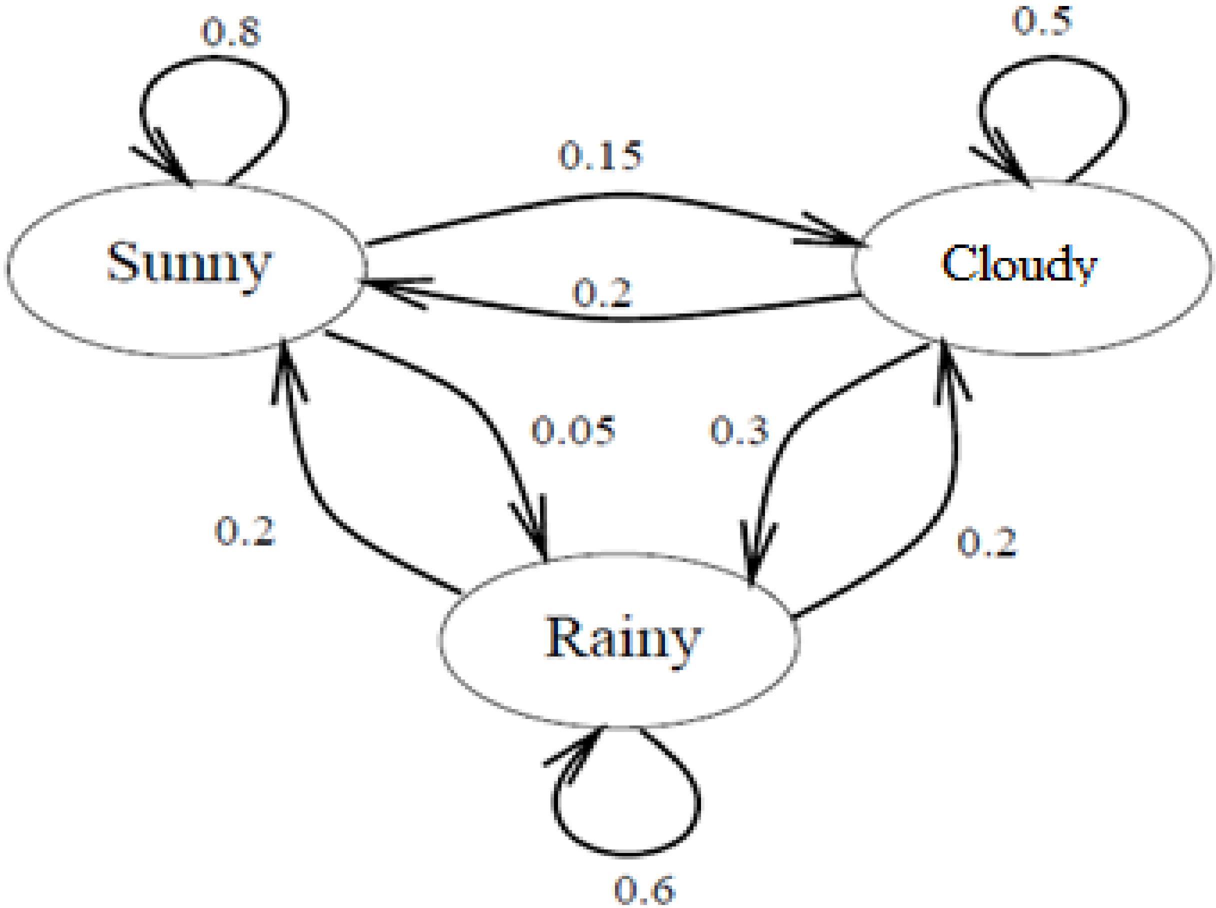
- A Markov Model is a stochastic model which models temporal or sequential data, i.e., data that are ordered.
- It provides a way to model the dependencies of current information (e.g. weather) with previous information.
- It is composed of states, transition scheme between states, and emission of outputs (discrete or continuous).
- Several goals can be accomplished by using Markov models:
  - Learn statistics of sequential data.
  - Do prediction or estimation.
  - Recognize patterns.

WHAT IS A HIDDEN MARKOV MODEL (HMM)?

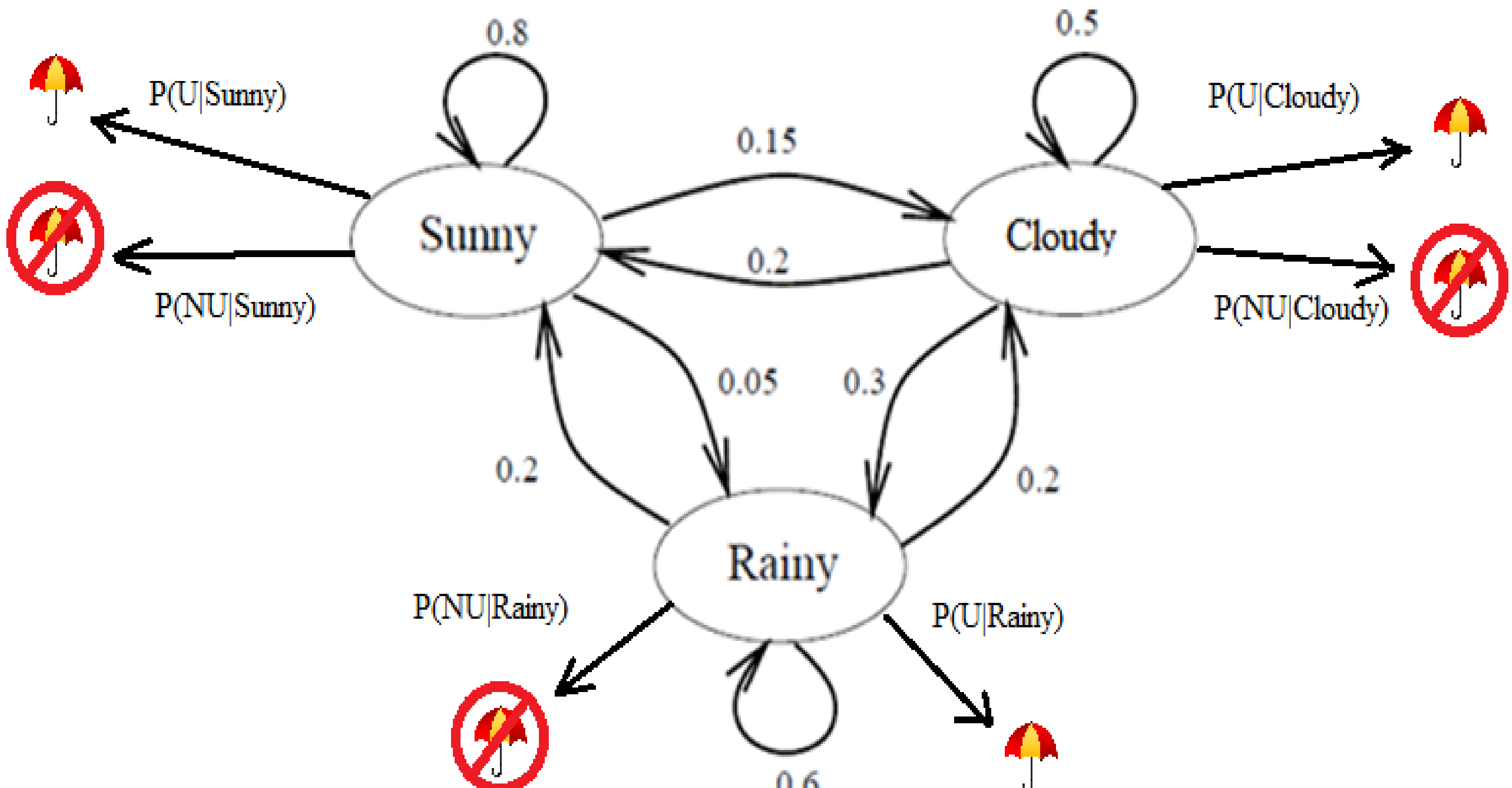
- A Hidden Markov Model, is a stochastic model where the states of the model are hidden. Each state can emit an output which is observed.
- **Imagine:** You were locked in a room for several days and you were asked about the weather outside. The only piece of evidence you have is whether the person who comes into the room bringing your daily meal is carrying an umbrella or not.
  - **What is hidden?** Sunny, Rainy, Cloudy
  - **What can you observe?** Umbrella or Not

Markov Chain vs. HMM

- Markov Chain:



- HMM:



U = Umbrella
NU = Not Umbrella

- Let's assume that t days had passed. Therefore, we will have an observation sequence $O = \{o_1, \dots, o_t\}$, where $o_i \epsilon \{Umbrella, Not\ Umbrella\}$.

- Each observation comes from an unknown state. Therefore, we will also have an unknown sequence $Q = \{q_1, \dots, q_t\}$, where $q_i \epsilon \{Sunny, Rainy, Cloudy\}$.

- We would like to know: $P(q_1, .., q_t | o_1, \dots, o_t)$.

HMM Mathematical Model

- From Bayes' Theorem, we can obtain the probability for a particular day as:

$$P(q_i|o_i) = \frac{P(o_i|q_i)P(q_i)}{P(o_i)}$$

For a sequence of length t:

$$P(q_1, \dots, q_t|o_1, \dots, o_t) = \frac{P(o_1, \dots, o_t|q_1, \dots, q_t)P(q_1, \dots, q_t)}{P(o_1, \dots, o_t)}$$

- From the Markov property:

$$P(q_1, \ldots, q_t) = \prod_{i=1}^{t} P(q_i | q_{i-1})$$

- Independent observations assumption:

$$P(o_1, \ldots, o_t | q_1, \ldots, q_t) = \prod_{i=1}^{t} P(o_i | q_i)$$

- Thus:

$$P(q_1, \dots, q_t | o_1, \dots, o_t) \propto \underbrace{\prod_{i=1}^{t} P(o_i|q_i) \prod_{i=1}^{t} P(q_i|q_{i-1})}$$

HMM Parameters:

- Transition probabilities $P(q_i|q_{i-1})$
- Emission probabilities $P(o_i|q_i)$
- Initial state probabilities $P(q_i)$

HMM PARAMETERS

- A HMM is governed by the following parameters:

$$\lambda = \{A, B, \pi\}$$

  - State-transition probability matrix A
  - Emission/Observation/State Conditional Output probabilities B
  - Initial (prior) state probabilities π

- Determine the fixed number of states (N):

$$S = \{s_1, \ldots, s_N\}$$

- State-transition probability matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & . & . & . & a_{1N} \\ a_{21} & a_{23} & . & . & . & a_{2N} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ a_{N1} & a_{N2} & . & . & . & a_{NN} \end{bmatrix}$$

$\sum_{j=1}^{N} a_{ij} = 1$ (Each row/Outgoing arrows)

$a_{ij} = P(q_t = s_j \mid q_{t-1} = s_i), \quad 1 \leq i, j \leq N$

$a_{ij} \geq 0$

$a_{ij} \rightarrow Transisiton\ probability\ from\ state\ s_i\ to\ s_j$

$$s_i \xrightarrow{a_{ij}} s_j$$

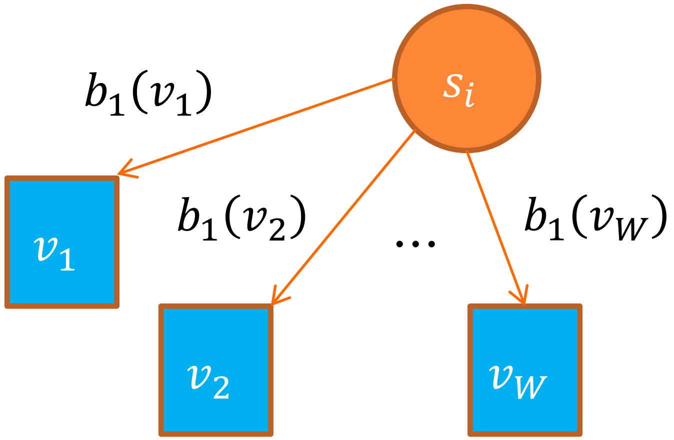
- Emission probabilities: A state will generate an observation (output), but a decision must be taken according on how to model the output, i.e., as discrete or continuous.
  - **Discrete outputs are modeled using pmfs.**
  - Continuous outputs are modeled using pdfs.

- Discrete Emission Probabilities:

Observation Set: $V = \{v_1, \dots, v_W\}$

$$b_i(v_k) = P(o_t = v_k | q_t = s_i), \qquad 1 \le k \le W$$

$$B = \begin{bmatrix} b_1(v_1) & b_1(v_2) & \cdot & \cdot & \cdot & b_1(v_W) \\ b_2(v_1) & b_2(v_2) & \cdot & \cdot & \cdot & b_2(v_W) \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ b_N(v_1) & b_N(v_2) & \cdot & \cdot & \cdot & b_N(v_W) \end{bmatrix}$$



- Initial (prior) probabilities: these are the probabilities of starting the observation sequence in state q_i.

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ . \\ . \\ . \\ \pi_N \end{bmatrix}$$

$$\pi_i = P(q_1 = s_i), \qquad 1 \le i \le N$$

$$\sum_{i=1}^{N} \pi_i = 1$$

HMM Example: Coins & Dice

$P(H|Red\ Coin) = 0.9$

$P(T|Red\ Coin) = 0.1$

$P(H|Green\ Coin) = 0.95$

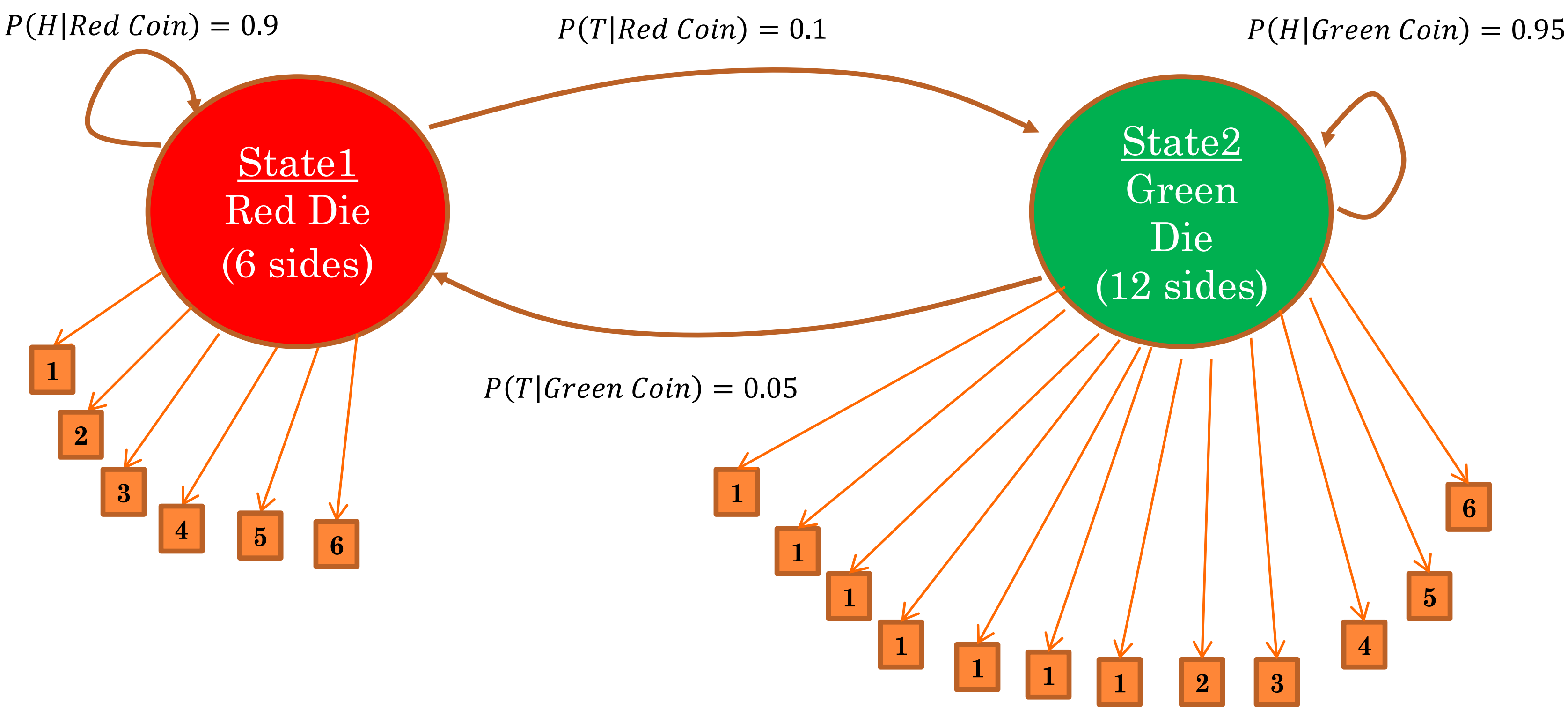
$P(T|Green\ Coin) = 0.05$

$Outputs = \{1, 2, 3, 4, 5, 6\}$

$Outputs = \{1, 1, 1, 1, 1, 1, 1, 2, 3, 4, 5, 6\}$

http://www.mathworks.com/help/stats/hidden-markov-models-hmm.html

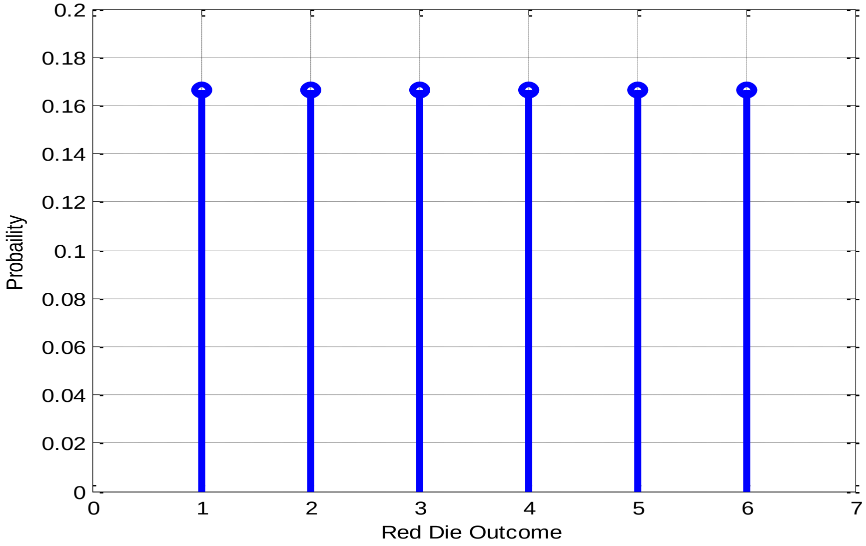
HMM Example: Coins & Dice



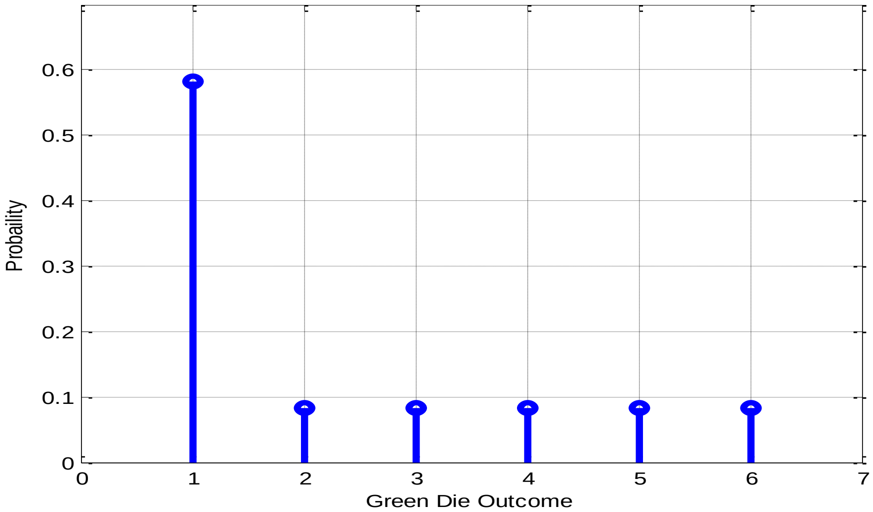
$$A = \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix} \qquad \pi = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

http://www.mathworks.com/help/stats/hidden-markov-models-hmm.html

HMM Example: Coins & Dice

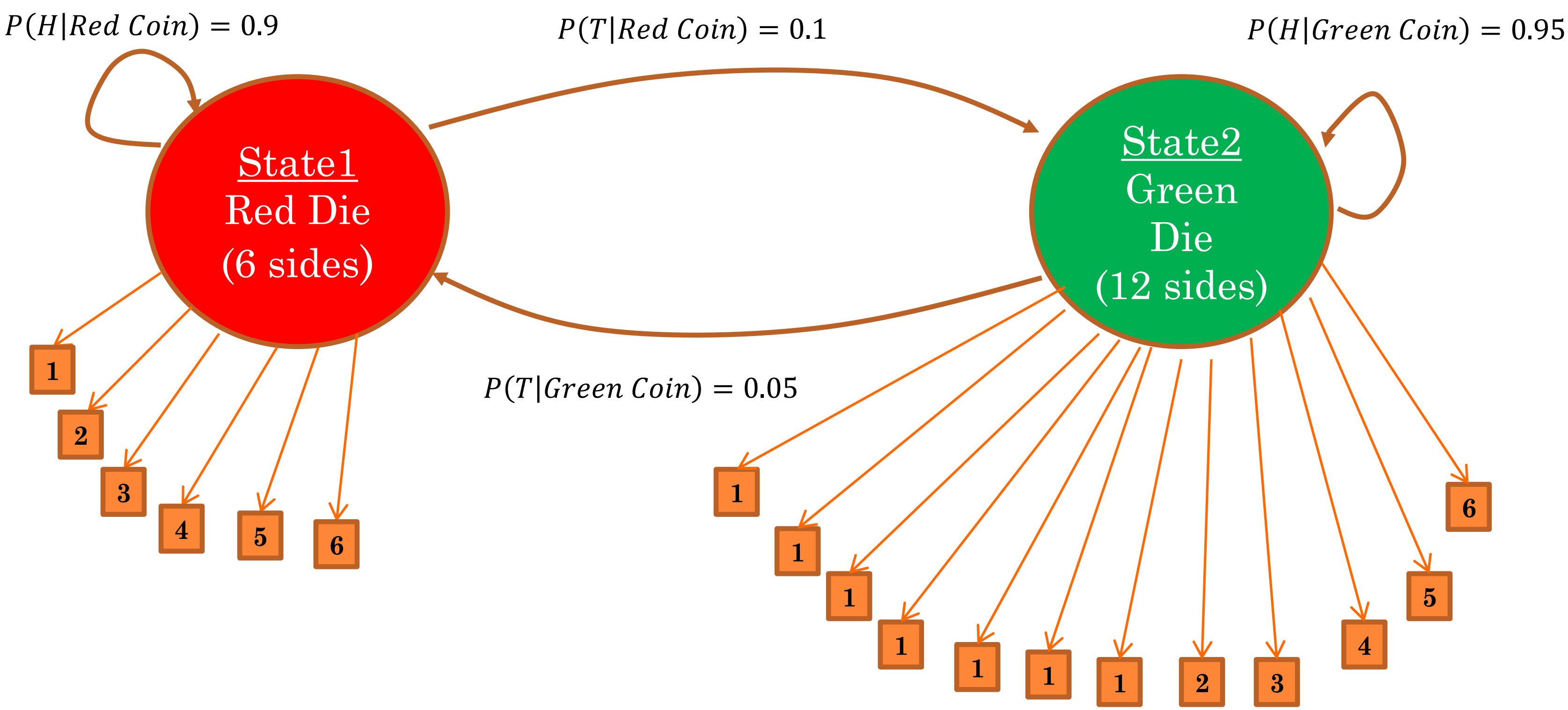


$$b_1(o_t) = \{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$$



$$b_2(o_t) = \{\frac{7}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}\}$$

HMM Example: Coins & Dice



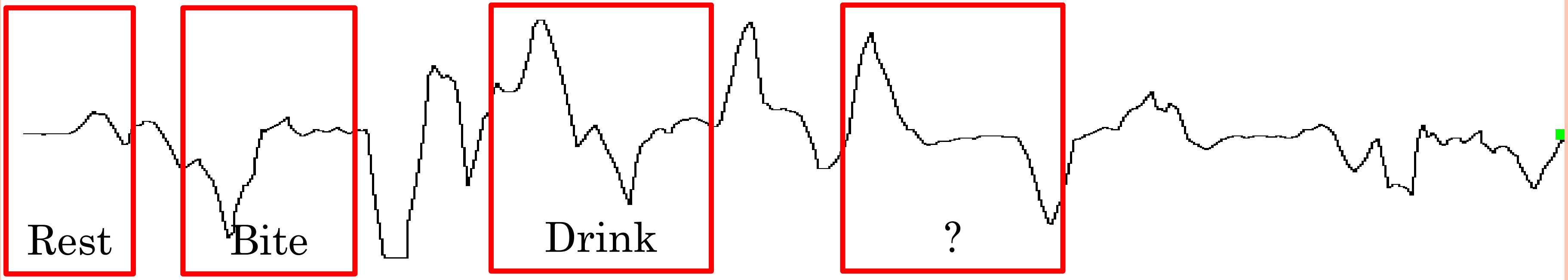
$$A = \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix} \qquad \pi = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad B = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{7}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \end{bmatrix}$$

http://www.mathworks.com/help/stats/hidden-markov-models-hmm.html

HMM to Classify Wrist Motions related to Eating Activities

273 Participants

Wrist Motion:



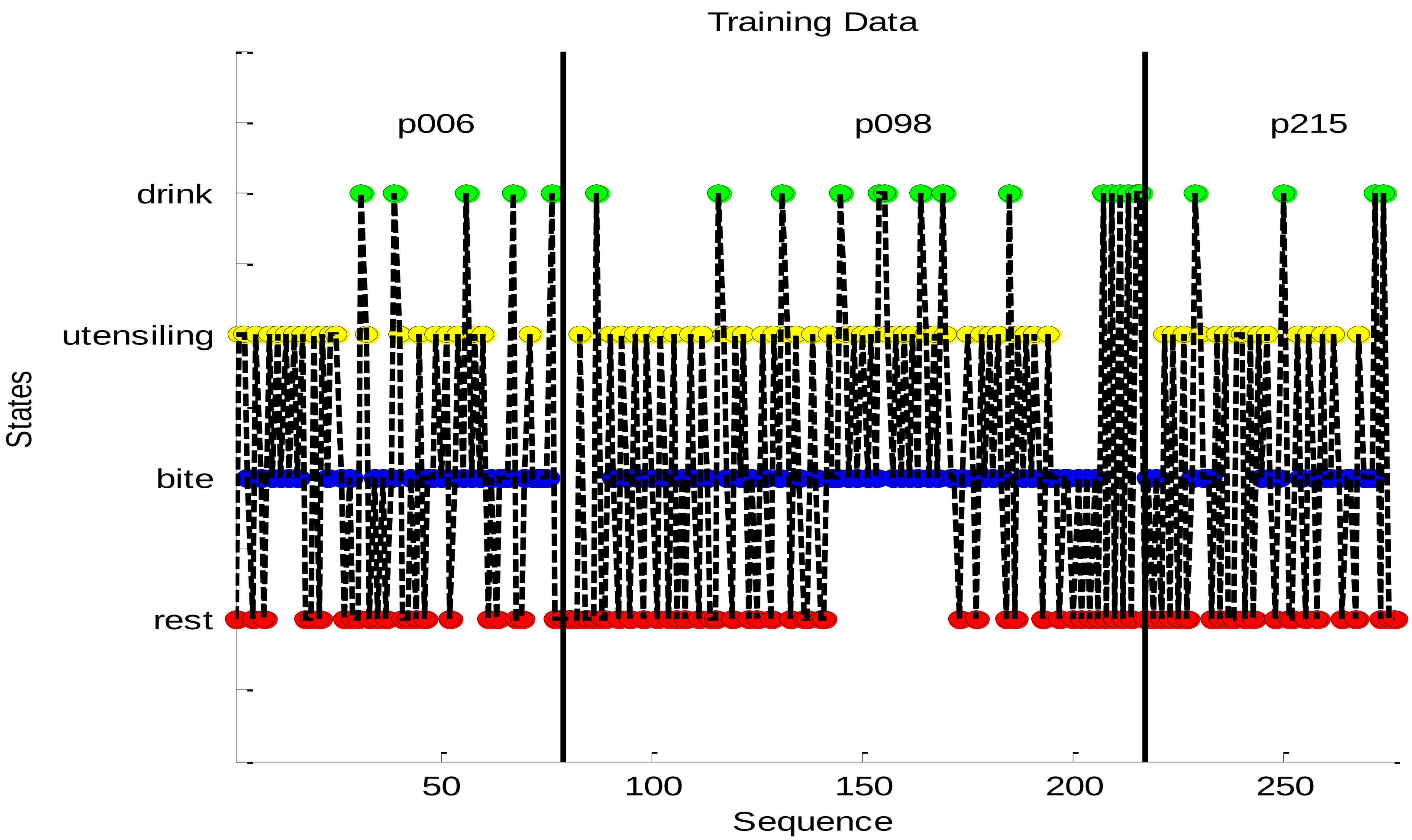
The "Language"

Word	Definition
Bite	When hand's motion towards the mouth starts until hand's motion away from the mouth ends (cannot occur than once a second)
Rest	The hand is not engaged in any discernible action.
Utensiling	Using utensil(s)/instrument to manipulate, stir, mix or prepare food(s) for consumption (boundaries defined by discernible intent).
Drink	When hand's motion towards mouth starts with intent to drink until hand's motion way from mouth ends and intent to drink has clearly ended.

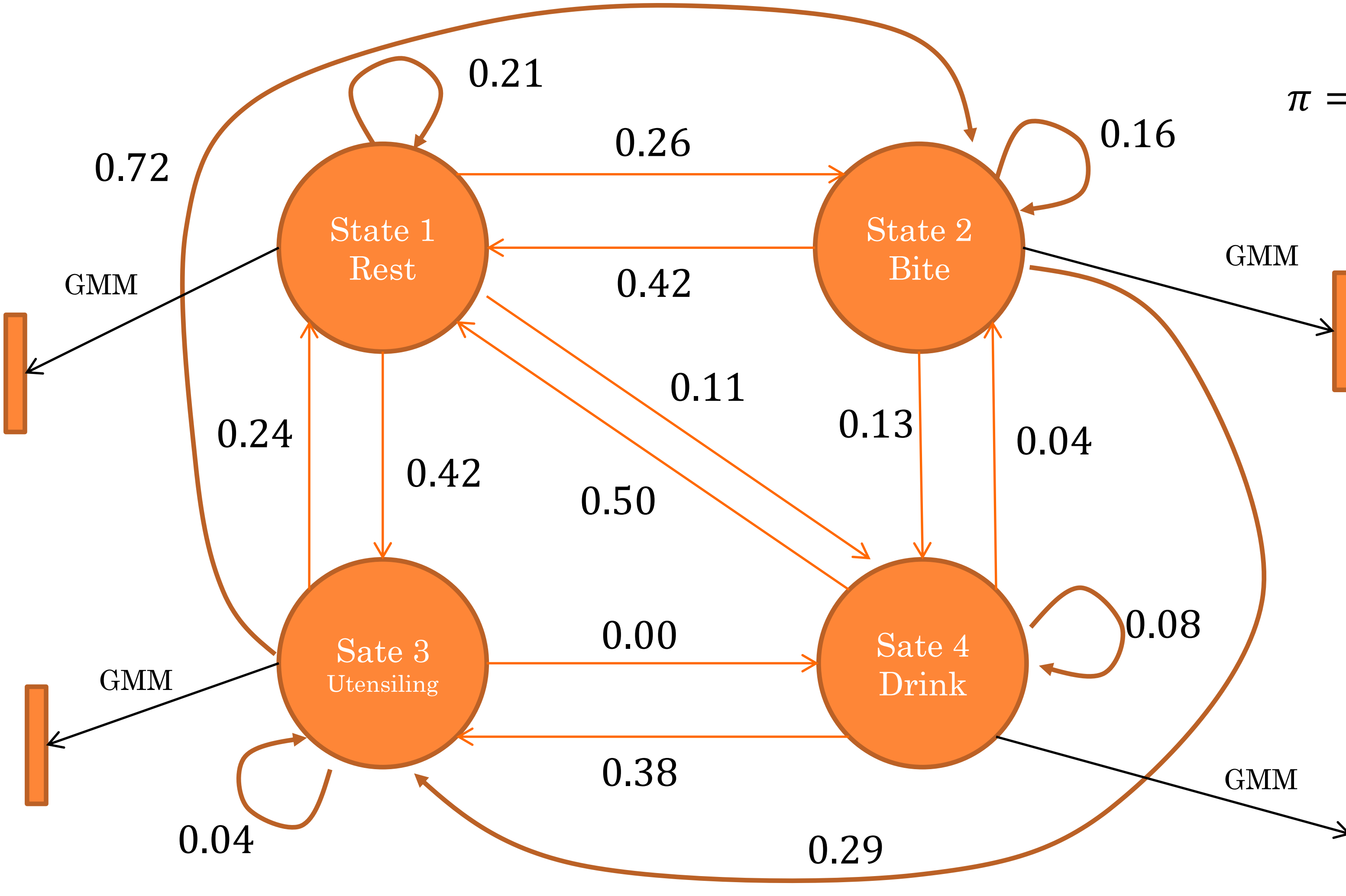
Data

Word	p006	p098	p215	Total
Rest	24	44	21	87
Utensiling	21	37	16	74
Bite	29	44	18	91
Drink	5	15	4	24

Data Sequence:



A, π:



$$\pi = \begin{bmatrix} 0.32 \\ 0.33 \\ 0.26 \\ 0.09 \end{bmatrix}$$

What can we do next?

- **State Sequence Decoding (Viterbi Algorithm):** Given a HMM we can find the best single state sequence (path) $Q = q_1, \ldots, q_T$ that best explains a known observation sequence $O = o_1, \ldots, o_T$.

- **Observation Sequence Evaluation (Forward-Backward Algorithm):** Evaluate a sequence of observations $O = o_1, \ldots, o_T$ given several alternative HMMs, and determine which one best recognizes the observation sequence (classification).

References

- Rabiner, L.R.; , **"A tutorial on hidden Markov models and selected applications in speech recognition,"** *Proceedings of the IEEE* , vol.77, no.2, pp.257-286, Feb 1989
- John R. Deller, John, and John H. L. Hansen. **“Discrete-Time Processing of Speech Signals”.** Prentice Hall, New Jersey, 1987.
- Barbara Resch (modified Erhard and Car Line Rank and Mathew Magimai-doss); **“Hidden Markov Models A Tutorial for the Course Computational Intelligence.”**
- Henry Stark and John W. Woods. **“Probability and Random Processes with Applications to Signal Processing (3rd Edition).”** Prentice Hall, 3 edition, August 2001.
- HTKBook: http://htk.eng.cam.ac.uk/docs/docs.shtml