# Hidden Markov Model (Lecture-3)

Anil Sharma

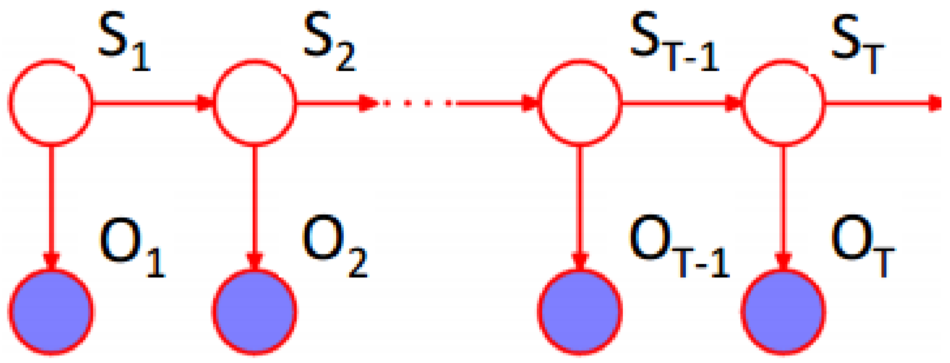
INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

# Hidden Markov Model (changed notations)

- Parameters – stationary/homogeneous markov model (independent of time t)

Initial probabilities

$p(S_1 = i) = \pi_i$




Transition probabilities

$p(S_t = j | S_{t-1} = i) = p_{ij}$

$$p(\{S_t\}_{t=1}^T, \{O_t\}_{t=1}^T) = p(S_1) \prod_{t=2}^T p(S_t|S_{t-1}) \prod_{t=1}^T p(O_t|S_t)$$

Emission probabilities

$p(O_t = y | S_t = i) = q_i^y$

# Three Basic Problems in HMM

- Evaluation – What is the probability of the observed sequence? - **Forward Algorithm**
- Decoding – What is the probability that the third roll was loaded given the observed sequence? - **Forward-Backward Algorithm**

– What is the most likely die sequence given the observed sequence? - **Viterbi Algorithm**

- Learning – Under what parameterization is the observed sequence most probable? **Baum-Welch Algorithm (EM)**

# Three Basic Problems in HMM

- Evaluation – Given HMM parameters & observation seqn $\{O_t\}_{t=1}^T$

  find $p(\{O_t\}_{t=1}^T|\theta)$ prob of observed sequence

- Decoding – Given HMM parameters & observation seqn $\{O_t\}_{t=1}^T$

  find $\arg\max_{s_1,\ldots,s_T} p(\{S_t\}_{t=1}^T|\{O_t\}_{t=1}^T,\theta)$ most probable sequence of hidden states

- Learning – Given HMM with unknown parameters and $\{O_t\}_{t=1}^T$ observation sequence

  find $\arg\max_\theta p(\{O_t\}_{t=1}^T|\theta)$ parameters that maximize likelihood of observed data

# Queries

**GIVEN:** A sequence of rolls by the casino player

1245526462146146136136661664661636616366163616515615115146123562344

**QUESTION**

- How likely is this sequence, given our model of how the casino works?
  - This is the **EVALUATION** problem in HMMs

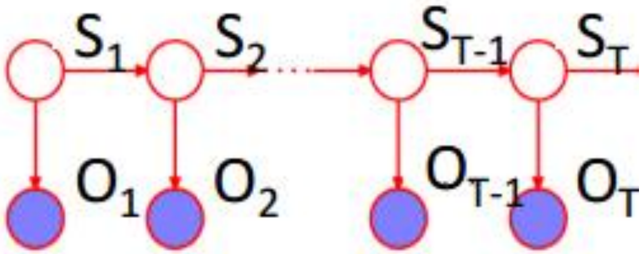
- What portion of the sequence was generated with the fair die, and what portion with the loaded die?
  - This is the **DECODING** question in HMMs

- How “loaded” is the loaded die? How “fair” is the fair die? How often does the casino player change from fair to loaded, and back?
  - This is the **LEARNING** question in HMMs

# 1. Evaluation Problem

- Given HMM parameters $p(S_1), p(S_t|S_{t-1}), p(O_t|S_t)$ & observation sequence $\{O_t\}_{t=1}^T$

find probability of observed sequence




$$
\begin{aligned}
p(\{O_t\}_{t=1}^T) &= \sum_{S_1,\ldots,S_T} p(\{O_t\}_{t=1}^T, \{S_t\}_{t=1}^T) \\
&= \sum_{S_1,\ldots,S_T} p(S_1) \prod_{t=2}^{T} p(S_t|S_{t-1}) \prod_{t=1}^{T} p(O_t|S_t)
\end{aligned}
$$

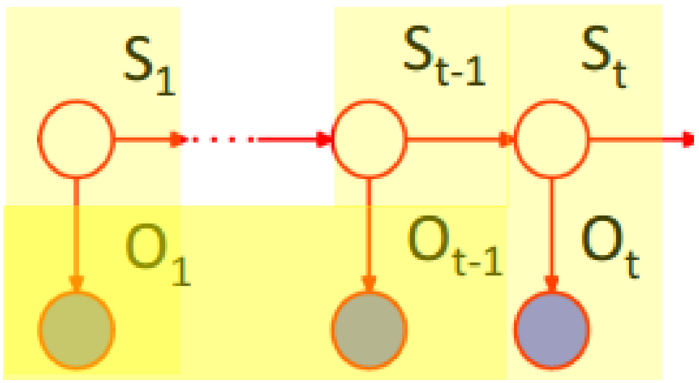
requires summing over all possible hidden state values at all times – $K^T$ exponential # terms!

Instead: $p(\{O_t\}_{t=1}^T) = \sum_k \underbrace{p(\{O_t\}_{t=1}^T, S_T = k)}_{\alpha_T^k}$ Compute recursively

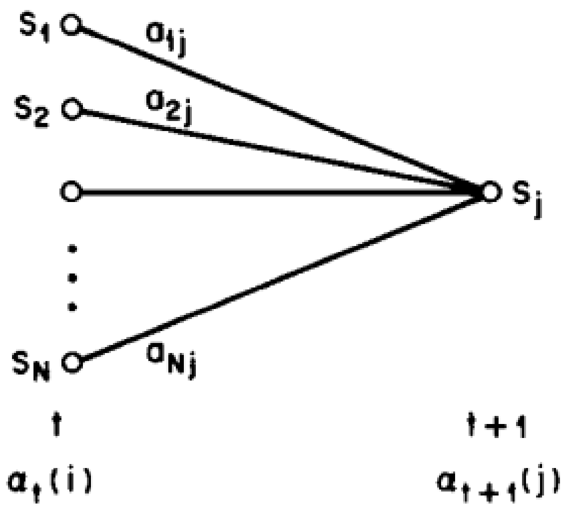
# Forward Probability

$$p(\{O_t\}_{t=1}^{T}) = \sum_k p(\{O_t\}_{t=1}^{T}, S_T = k) = \sum_k \alpha_T^k$$

Compute forward probability $\alpha_t^k$ recursively over t




$$\alpha_t^k \quad := \quad p(O_1, \ldots, O_t, S_t = k)$$



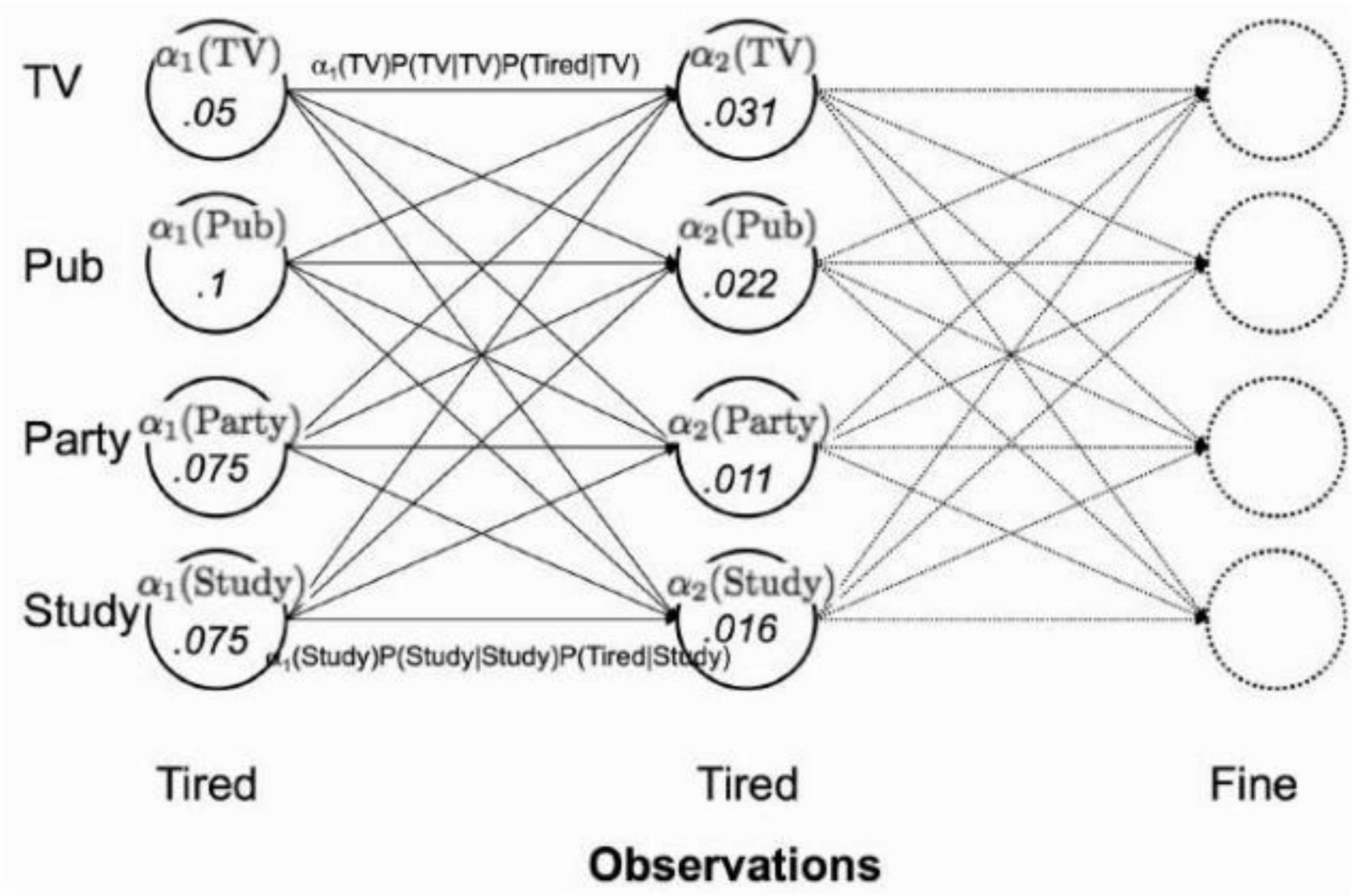

Introduce $S_{t-1}$

Chain rule

Markov assumption

$$= \quad p(O_t|S_t = k) \sum_i \alpha_{t-1}^i p(S_t = k|S_{t-1} = i)$$

# Forward variables (Graphically)



# Forward Algorithm

Can compute $\alpha_t^k$ for all k, t using dynamic programming:

- Initialize: $\alpha_1^k = p(O_1 | S_1 = k)\, p(S_1 = k)$ for all k

- Iterate: for t = 2, ..., T

  $\alpha_t^k = p(O_t | S_t = k) \sum_i \alpha_{t-1}^i\, p(S_t = k | S_{t-1} = i)$ for all k

- Termination: $p(\{O_t\}_{t=1}^T) = \sum_k \alpha_T^k$

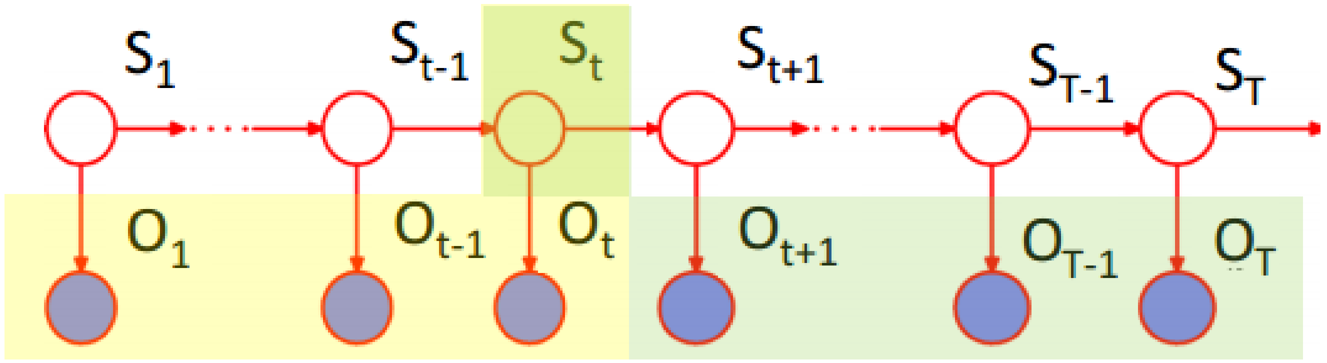
# 2. Decoding Problem -1

- Given HMM parameters $p(S_1), p(S_t|S_{t-1}), p(O_t|S_t)$ & observation sequence $\{O_t\}_{t=1}^T$

  find probability that hidden state at time t was k $p(S_t = k|\{O_t\}_{t=1}^T)$

$$
\begin{aligned}
p(S_t = k, \{O_t\}_{t=1}^T) &= p(O_1, \ldots, O_t, S_t = k, O_{t+1}, \ldots, O_T) \\
&= \underbrace{p(O_1, \ldots, O_t, S_t = k)}_{\alpha_t^k}\underbrace{p(O_{t+1}, \ldots, O_T|S_t = k)}_{\beta_t^k}
\end{aligned}
$$

Compute recursively

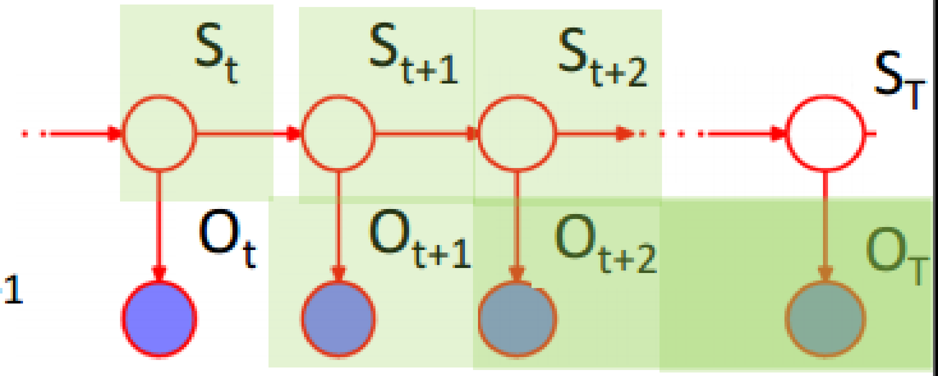


# Backward Probability

$$p(S_t = k, \{O_t\}_{t=1}^{T}) = p(O_1, \ldots, O_t, S_t = k)p(O_{t+1}, \ldots, O_T | S_t = k) = \alpha_t^k \beta_t^k$$

Compute forward probability $\beta_t^k$ recursively over t

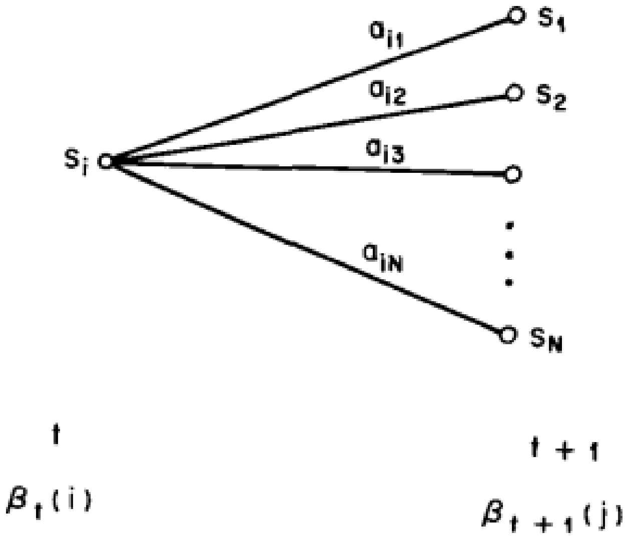
$$\beta_t^k \quad := \quad p(O_{t+1}, \ldots, O_T | S_t = k)$$




Introduce $S_{t+1}$

Chain rule

Markov assumption




$$= \sum_i p(S_{t+1} = i | S_t = k)p(O_{t+1} | S_{t+1} = i)\beta_{t+1}^i$$

# Forward-Backward Algorithm

Can compute $\beta_t^k$ for all k, t using dynamic programming:

- Initialize: $\beta_T^k = 1$ for all k

- Iterate: for t = T-1, …, 1

$$\beta_t^k = \sum_i p(S_{t+1} = i | S_t = k) p(O_{t+1} | S_{t+1} = i) \beta_{t+1}^i \quad \text{for all k}$$

- Termination: $p(S_t = k, \{O_t\}_{t=1}^T) = \alpha_t^k \beta_t^k$

$$p(S_t = k | \{O_t\}_{t=1}^T) = \frac{p(S_t = k, \{O_t\}_{t=1}^T)}{p(\{O_t\}_{t=1}^T)} = \frac{\alpha_t^k \beta_t^k}{\sum_i \alpha_t^i \beta_t^i}$$

# Most likely state & most likely sequence

- Most likely state assignment at time t

$$\arg\max_k p(S_t = k | \{O_t\}_{t=1}^T) = \arg\max_k \alpha_t^k \beta_t^k$$

E.g. Which die was most likely used by the casino in the third roll given the observed sequence?

- Most likely assignment of state sequence

$$\arg \max_{\{S_t\}_{t=1}^T} p(\{S_t\}_{t=1}^T | \{O_t\}_{t=1}^T)$$

E.g. What was the most likely sequence of die rolls used by the casino given the observed sequence?

**Not the same solution !**

MLA of x?
MLA of (x,y)?

| x | y | P(x,y) |
|---|---|---|
| 0 | 0 | 0.35 |
| 0 | 1 | 0.05 |
| 1 | 0 | 0.3 |
| 1 | 1 | 0.3 |

# Contd.. Decoding Problem -2

- Given HMM parameters $p(S_1), p(S_t|S_{t-1}), p(O_t|S_t)$ & observation sequence $\{O_t\}_{t=1}^T$

  find most likely assignment of state sequence

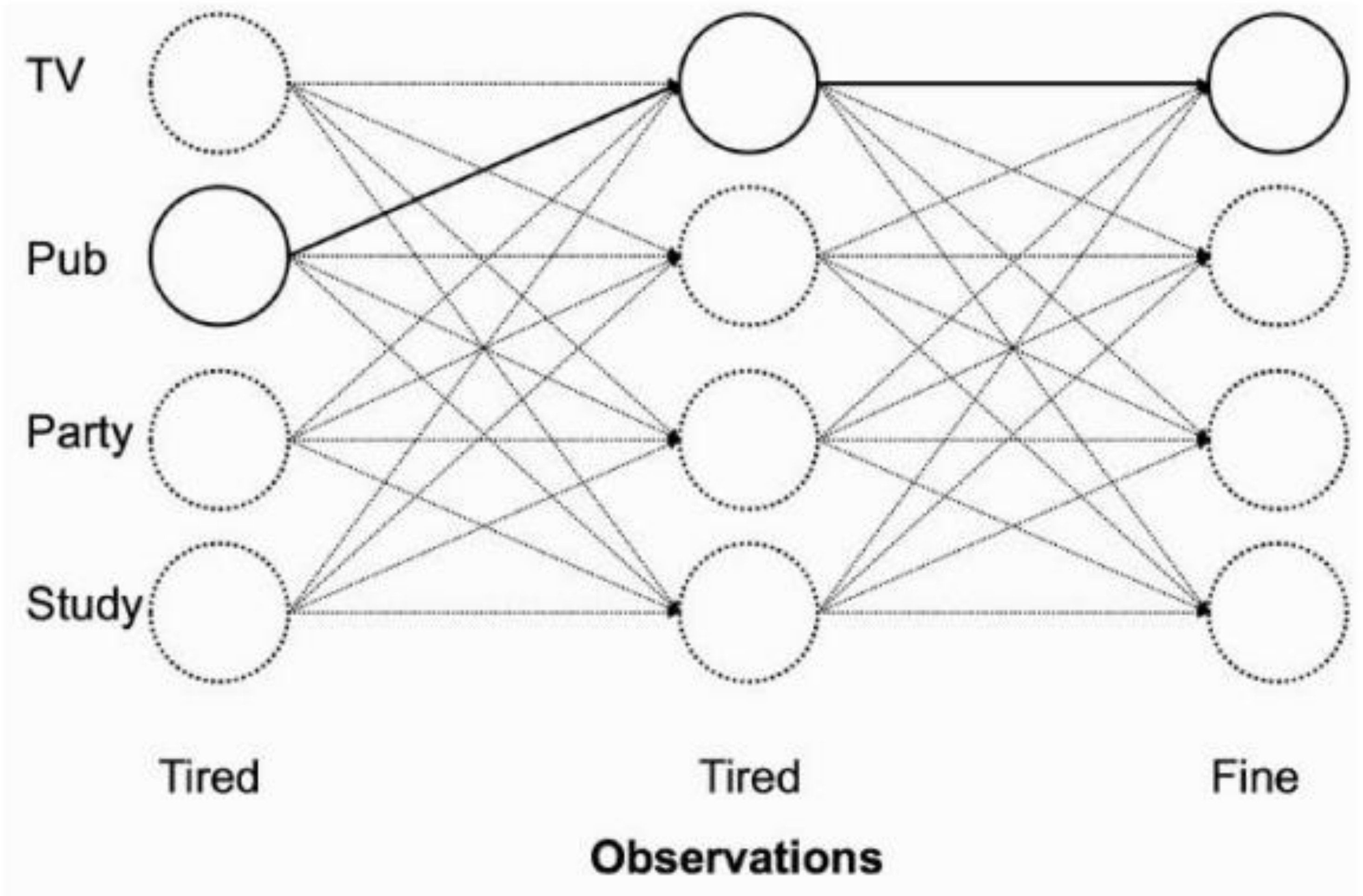
$$\arg\max_{\{S_t\}_{t=1}^T} p(\{S_t\}_{t=1}^T | \{O_t\}_{t=1}^T) = \arg\max_{\{S_t\}_{t=1}^T} p(\{S_t\}_{t=1}^T, \{O_t\}_{t=1}^T)$$

$$= \arg\max_k \underbrace{\max_{\{S_t\}_{t=1}^{T-1}} p(S_T = k, \{S_t\}_{t=1}^{T-1}, \{O_t\}_{t=1}^T)}_{V_T^k}$$

Compute recursively

$V_T^k$ - probability of most likely sequence of states ending at state $S_T = k$

# Graphically



# Viterbi Algorithm

$$\max_{\{S_t\}_{t=1}^T} p(\{S_t\}_{t=1}^T, \{O_t\}_{t=1}^T) = \max_k V_T^k$$

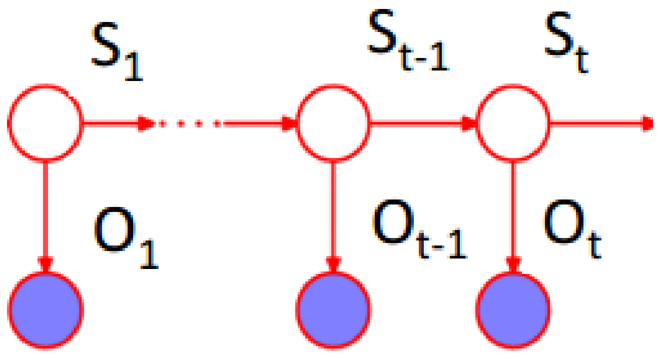
Compute probability $V_t^k$ recursively over t

$$V_t^k \quad := \quad \max_{S_1,\ldots,S_{t-1}} p(S_t = k, S_1, \ldots, S_{t-1}, O_1, \ldots, O_t)$$

$$\vdots$$

Bayes rule

Markov assumption




$$= \quad p(O_t|S_t = k) \max_i p(S_t = k|S_{t-1} = i) V_{t-1}^i$$

# Contd.. (Viterbi Algorithm)

Can compute $V_t^k$ for all k, t using dynamic programming:

- Initialize: $V_1^k = p(O_1|S_1=k)p(S_1 = k)$ for all k

- Iterate: for t = 2, ..., T

$$V_t^k = p(O_t|S_t = k) \max_i p(S_t = k|S_{t-1} = i) V_{t-1}^i$$ for all k

- Termination: $$\max_{\{S_t\}_{t=1}^T} p(\{S_t\}_{t=1}^T, \{O_t\}_{t=1}^T) = \max_k V_T^k$$

Traceback:

$$S_T^* = \arg\max_k V_T^k$$

$$S_{t-1}^* = \arg\max_i p(S_t^*|S_{t-1} = i) V_{t-1}^i$$

# What you should know

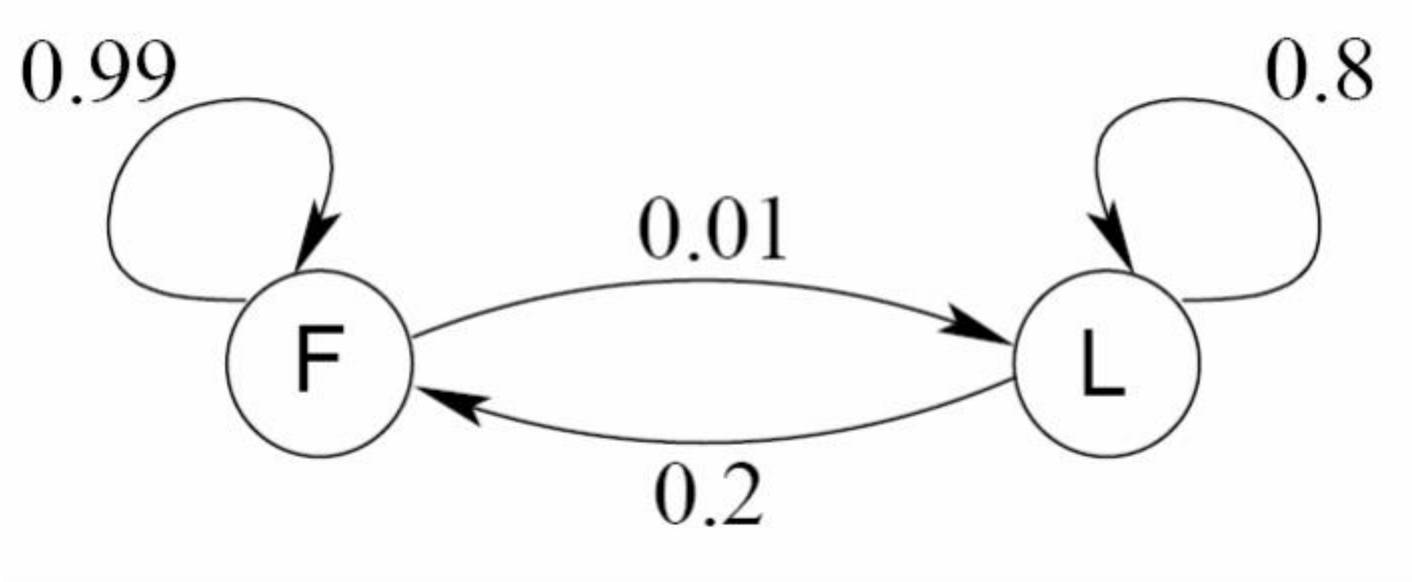
- Useful for modeling sequential data with few parameters using discrete hidden states that satisfy Markov assumption
- **Representation** - initial prob, transition prob, emission prob, State space representation
- **Algorithms for inference and learning in HMMs**

–Computing marginal likelihood of the observed sequence: forward algorithm

–Predicting a single hidden state: forward-backward

–Predicting an entire sequence of hidden states: Viterbi

# Example- Dishonest Casino




P(1|F) = 1/6
P(2|F) = 1/6
P(3|F) = 1/6
P(4|F) = 1/6
P(5|F) = 1/6
P(6|F) = 1/6

P(1|L) = 1/10
P(2|L) = 1/10
P(3|L) = 1/10
P(4|L) = 1/10
P(5|L) = 1/10
P(6|L) = 1/2

# Question

- Given your HMM model and observation sequence for three time steps, O = {6,2,6}. Find
  - The probability of the observation sequence {6,2,6}
  - The most likely state sequence.
- Hint: Use Forward algorithm and Viterbi algorithm

$$V_t^k = p(O_t|S_t = k) \max_i p(S_t = k|S_{t-1} = i) V_{t-1}^i$$

for all k

$$S_T^* = \arg\max_k V_T^k$$

$$S_{t-1}^* = \arg\max_i p(S_t^*|S_{t-1} = i) V_{t-1}^i$$

# Solution for $S^*$

| | 6 | 2 | 6 |
|---|---|---|---|
| F | $(1/6)\times(1/2)$ $= 1/12$ | $(1/6)\times\max\{(1/12)\times0.99,\ (1/4)\times0.2\}$ $= 0.01375$ | $(1/6)\times\max\{0.01375\times0.99,\ 0.02\times0.2\}$ $= 0.00226875$ |
| L | $(1/2)\times(1/2)$ $= 1/4$ | $(1/10)\times\max\{(1/12)\times0.01,\ (1/4)\times0.8\}$ $= 0.02$ | $(1/2)\times\max\{0.01375\times0.01,\ 0.02\times0.8\}$ $= 0.08$ |

- Most likely sequence after backtrack is $S^* = \{L,L,L\}$.

# Reference


- Rabiner, Lawrence R. "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE* 77.2 (1989): 257-286.
- Machine Learning: An algorithm perspective by Stephen Marsland
- Slides reference: http://www.cs.cmu.edu/~aarti/Class/10701_Spring14/slides/HMM.pdf

- Class notes are uploaded on backpack.