# Hidden Markov Models

Hosein Mohimani

GHC7717

hoseinm@andrew.cmu.edu

# Fair Bet Casino Problem

Dealer flips a coin and player bets on outcome

Dealer use either a fair coin (head and tail equally likely) or a biased coin (head with probability 3/4)

Can you guess which of the following sequences are generated by a fair coin, and which ones with a biased one ?

1 HTHHHHHTHTHTTTTHHTHTTTHHTTTHHHTHHTTTHTTT

2 THTHHHTHHHHHTHTTHHHTHHHHHHHHHHHTHHHHHHH

3 HHTHTTHHHTTHHTTTHHHHHHHTHTHHHTHHHHHHHHH

4 HTHTHTTTHTTHHHHTHHTHTTHHHTHTTTTHTHTHTHHH

# Fair Bet Casino Problem

Dealer flips a coin and player bets on outcome

Dealer use either a fair coin (head and tail equally likely) or a biased coin (head with probability 3/4)

1 HTHHHHHTHTHTTTTHHTHTTTHHTTTHHHTHHTTTHTTT 19/40

2 THTHHHTHHHHHTHTTHHHTHHHHHHHHHHHTHHHHHHH 31/40

3 HHTHTTHHHTTHHTTTHHHHHHHTHTHHHTHHHHHHHHH 28/40

4 HTHTHTTTHTTHHHHTHHTHTTHHHTHTTTTHTHTHTHHH 21/40

# Fair Bet Casino Problem

The dealer changes the coin occasionally

HTHTTHHTHTHTTTTHHTHTTTHHTTTHHHTHHTTTHTTTTHTHH
HTHHHHTHTTHHHTHHHHHHHHHHHTHHHHHHHHHHTHTTHHH
TTHHTTTHHHHHHHTHTHHHTHHHHHHHHHHHTHTHTTTHTTHHH
HTHHTHTTHHHTHTTTTHTHTHTHHH

Can you guess which parts the fair coin was used,
and in which parts the Biased coin was used ?

# Fair Bet Casino Problem

The dealer changes the coin occasionally

HTHTTHHTHTHTTTTHHTHTTTHHTTTHHHTHHTTTHTTTTHTHH
HTHHHHHTHTTHHHTHHHHHHHHHHHTHHHHHHHHHTHTTHHH
TTHHTTTHHHHHHHTHTHHHTHHHHHHHHHHTHTHTTTHTTHHH
HTHHTHTTHHHTHTTTTHTHTHTHHH

FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
FFBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBFFFFFFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF

# Fair Bet Casino Problem : multiple hypothesis testing

- Is the following sequence coming from a fair (1/2,1/2) coin, or a biased (3/4,1/4) coin ?

HTHHHHTHTHTTTTHHTHTTTHHTTTHHHTHHTTTHTTT    19/40

Hypothesis 1 : Fair coin (1/2,1/2)

Hypothesis 2 : Biased coin (3/4,1/4)

- Which of the two hypothesis is more likely to happen ?

# Lets compue the probabilities

- Consider sequence HTH
- What is the likelihood of HTH under fair coin hypothesis?
- What is the likelihood of HTH under biased coin hypothesis?
- Which of the two hypothesis is more likely ?

# Lets compute the probabilities

- $P(HTH|fair) = P(H|fair).P(T|fair).P(H|fair) = 1/2.1/2.1/2 = 1/8$
- $P(HTH|biased) = P(H|biased).P(T|biased).P(H|biased) = 3/4 \,.1/4 \,.\, 3/4 = 9/64$
- Biased coin hypothesis is slightly more likely

# Multiple Hypothesis Testing

*seq=HTHHHHHTHTHTTTTHHTHTTTHHTTTHHHTHHTTTHTTT*

- What is the likelihood of this sequence under each hypothesis?

Under fair hypothesis $\quad P(seq \mid fair) = (\frac{1}{2})^{40} = 9.10^{-13}$

Log-likelihood $\quad log\ P(seq \mid fair) = 40.log(1/2) = -12.04$

Under biased hypothesis $\quad P(seq \mid biased) = (\frac{3}{4})^{19} (\frac{1}{4})^{21} = 9.10^{-16}$

$log\ P(seq \mid biased) = 19.log(3/4) + 21.log(1/4) = -15.01$

- Fair hypothesis is more likely than biased hypothesis

# Conditional Probabilities

- $P(x|F) = \prod_{i=1}^{N} P^F(x_i) = \frac{1}{2^n}$
- $P(x|B) = \prod_{i=1}^{N} P^B(x_i) = \frac{1}{4^{n-k}} \frac{3^k}{4^k} = \frac{3^k}{4^n}$ ($k$ is the number of heads in $x$)
- $P(x|F) > P(x|B) \rightarrow$ dealer probably use a fair coin k $< \frac{n}{\log_2 3}$
- $P(x|F) < P(x|B) \rightarrow$ dealer probably use a biased coin k $> \frac{n}{\log_2 3}$

- Log odd-ratio : $\log_2 \frac{P(x|F)}{P(x|B)} = \sum_{i=1}^{n} \log_2 \frac{P^F(x_i)}{P^B(x_i)} = n - k \log_2 3$

# Markov Model of Fair Bet Casino

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F → | F → | F → | F → | F → | F → | B → | B → | B → | B → | B → | B → | B → | B → | F → |
| ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| H | H | T | T | H | T | H | H | T | H | H | H | H | T | T |

- We observe a sequence coin tosses HHTTHTHHTHHHHTT
- How can we predict when the coin was fair/biased ? e.g. the most likely sequence FFFFFFBBBBBBBBF that resulted in this outcome ?

# Fair Bet Casino Problem

- Given a sequence of coin tosses, determine when a dealer used a fair coin and when he used a biased coin

- Input : A sequence $x = x_1 x_2 \dots x_n$ of coin tosses (either H or T) made by two possible coins (F or B).

- Output : A sequence $\pi = \pi_1 \pi_2 \dots \pi_n$ with each $\pi_i$ being either $F$ or $B$ indicating that $x_i$ is the result of tossing the fair or biased coin, respectively.

# Is this even possible?

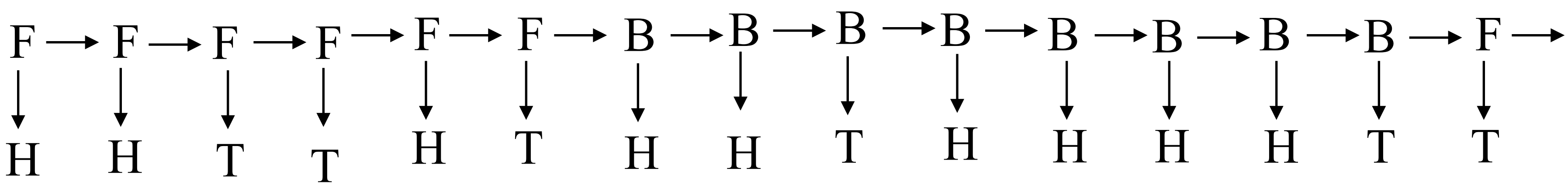
- In practice, any sequence $\pi = \pi_1\pi_2\ldots\pi_n$ can generate any outcome $x=x_1x_2\ldots x_n$

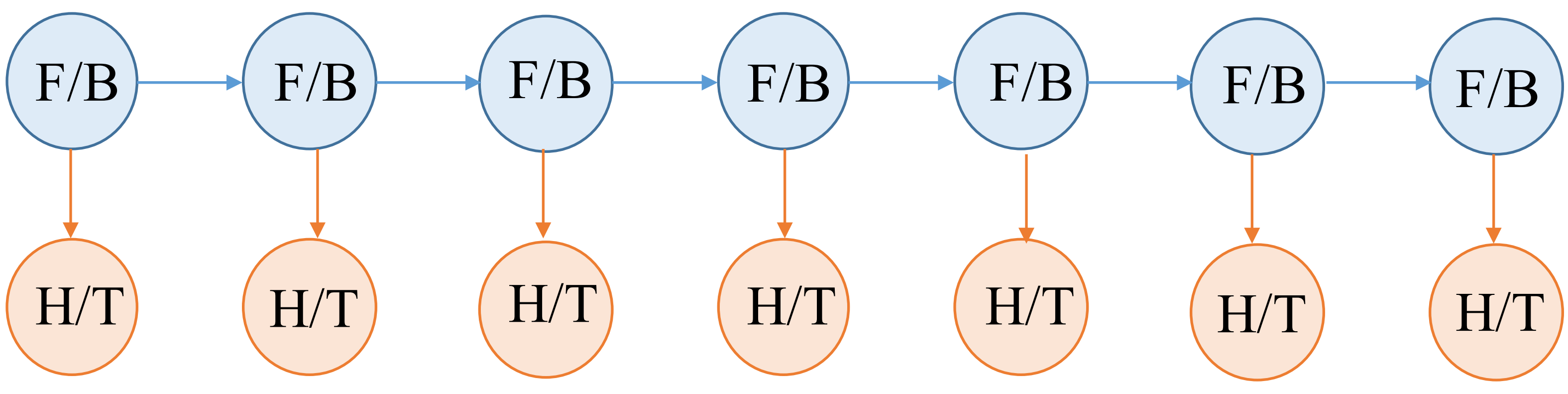
- We are interested in the most probable sequence $\pi$ that explains the outcome $x$

- Fair coin : $P^F(H)=P^F(T)=1/2$

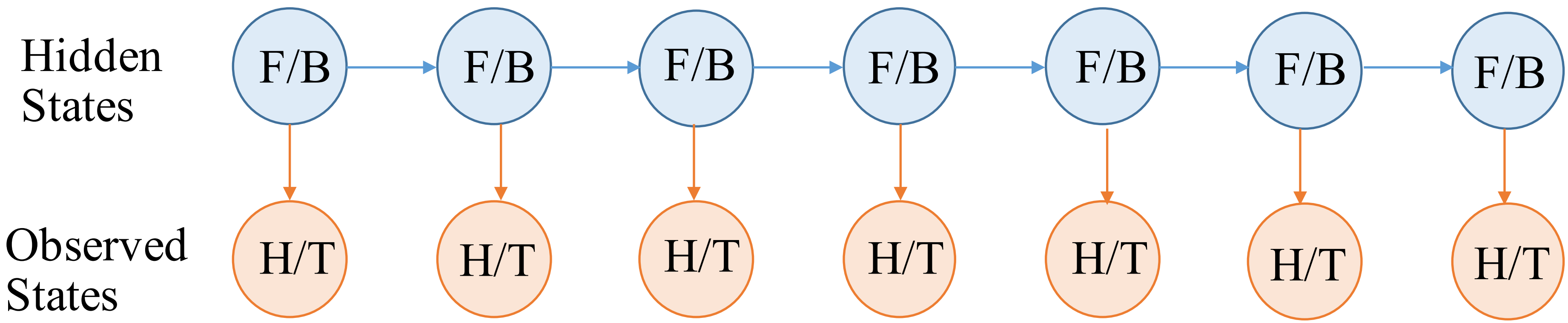
- Biased coin : $P^F(H)=3/4$, $P^F(T)=1/4$

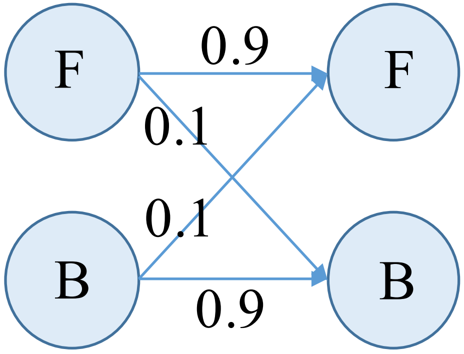
# Markov Model of Fair Bet Casino



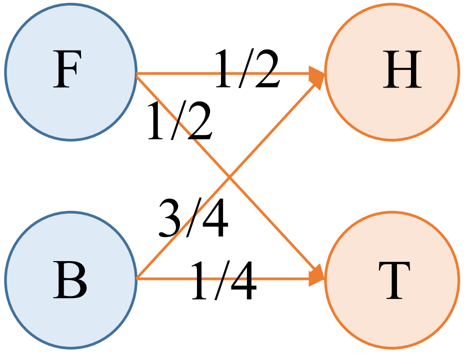



# Hidden Markov Model of Fair Bet Casino




Transition
Probabilities




Emission
Probabilities

# Hidden Markov Model

- An abstract machine with ability to produce output by coin tossing
- At the beginning, the machine is in one of the *k* hidden states (Fair or Biased in case of Fair Bet Casino)
- At each step, HMM makes two decisions :
  - What state will I move into (Fair or Biased)
  - What symbol from an alphabet $\Sigma$ will I emit (Head or Tail)
- The observer see emitted symbols, but not HMM states
- The goal of the observer is to infer the most likely state by analyzing the emitted symbols.

# HMM : formal definition

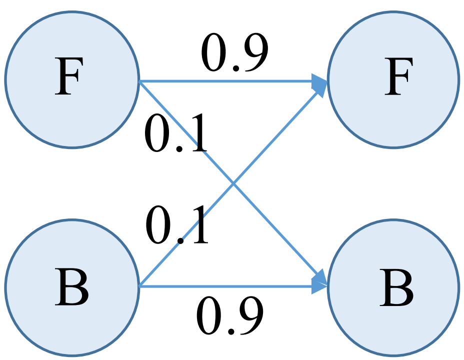
Formally, an HMM $\mathcal{M}$ is defined by

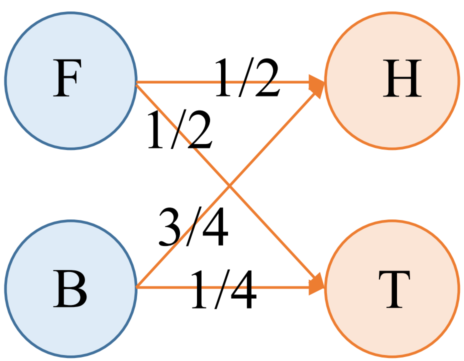
- An alphabet of emitted symbols $\Sigma$
- A set of hidden states $Q$
- A $|Q| x |Q|$ matrix of transition probabilities $A=(a_{kl})$, where $a_{kl}$ is the probability of changing to state $k$ after being in state $l$

A $|Q| x |\Sigma|$ matrix of emission probabilities $E=e_k(b)$, describing the probability of emitting the symbol $b$ during a step in which the HMM is in step $k$

# Back to the casino fair bet problem

- Each row of the matrix $A$ is a state die with $|Q|$ sides
- Each row of the matrix $E$ is a state die with $|\Sigma|$ sides
- The casino fair bet problem $\mathcal{M}(\Sigma,Q,A,E)$
  - $\Sigma = \{0,1\}$, corresponding to tails (0) or heads (1)
  - $Q = \{F,B\}$, corresponding to fair (F) and biased (B) coin
  - $a_{FF}=a_{BB}=0.9,\ a_{FB}=a_{BF}=0.1$
  - $e_F(0)=1/2,\ e_F(1)=1/2,\ e_B(0)=1/4,\ e_B(1)=3/4,$



# Casino Fair Bet problem : example

$$
\begin{array}{c} x \\ \pi \\ P(x_i|\pi_i) \\ P(\pi_{i-1} \to \pi_i) \end{array}
\quad = \quad
\left(\begin{array}{ccccccccccc}
0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\
F & F & F & B & B & B & B & B & F & F & F \\
\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{3}{4} & \frac{3}{4} & \frac{3}{4} & \frac{1}{4} & \frac{3}{4} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & \frac{9}{10} & \frac{9}{10} & \frac{1}{10} & \frac{9}{10} & \frac{9}{10} & \frac{9}{10} & \frac{9}{10} & \frac{1}{10} & \frac{9}{10} & \frac{9}{10}
\end{array}\right)
$$

$$
P(\pi_0 \to \pi_1) \cdot \prod_{i=1}^{n} P(x_i|\pi_i) P(\pi_i \to \pi_{i+1}) = a_{\pi_0,\pi_1} \cdot \prod_{i=1}^{n} e_{\pi_i}(x_i) \cdot a_{\pi_i,\pi_{i+1}} .
$$

$$
\left(\tfrac{1}{2} \cdot \tfrac{1}{2}\right)\left(\tfrac{1}{2} \cdot \tfrac{9}{10}\right)\left(\tfrac{1}{2} \cdot \tfrac{9}{10}\right)\left(\tfrac{3}{4} \cdot \tfrac{1}{10}\right)\left(\tfrac{3}{4} \cdot \tfrac{9}{10}\right)\left(\tfrac{3}{4} \cdot \tfrac{9}{10}\right)\left(\tfrac{1}{4} \cdot \tfrac{9}{10}\right)\left(\tfrac{3}{4} \cdot \tfrac{9}{10}\right)\left(\tfrac{1}{2} \cdot \tfrac{1}{10}\right)\left(\tfrac{1}{2} \cdot \tfrac{9}{10}\right)\left(\tfrac{1}{2} \cdot \tfrac{9}{10}\right)
$$

# Decoding Problem

- Goal : Find an optimal hidden path of states given observations.

- Input : Sequence of observations $x=x_1 \, . \, . \, . \, x_n$ generated by an HMM $\mathcal{M}(P, Q, A, \Sigma)$.

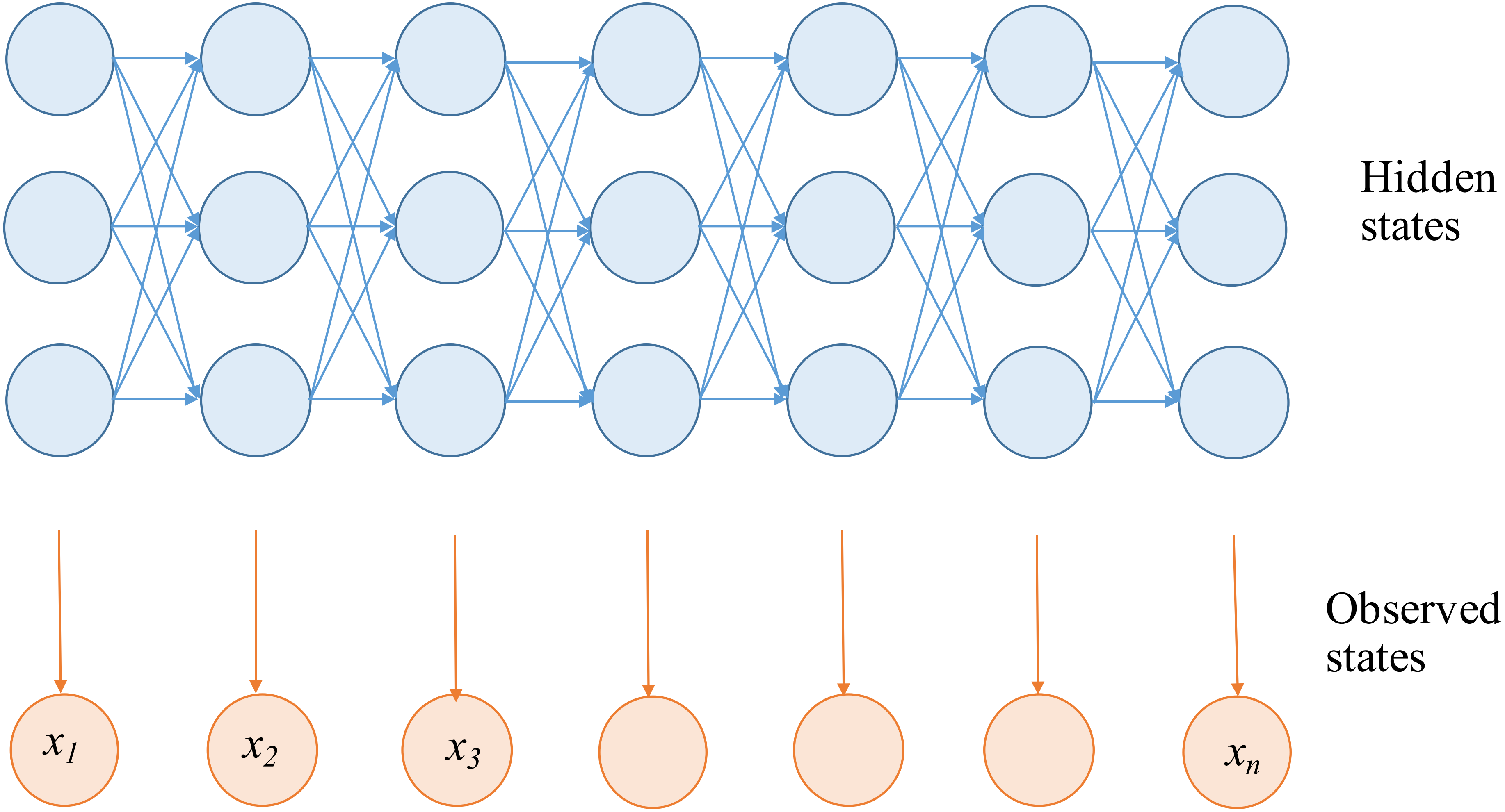
- Output: A path that maximizes $P(x|\pi)$ over all possible paths

$\pi=\pi_1 \, \pi_2 \ldots \pi_n$.

# Viterbi Decoding

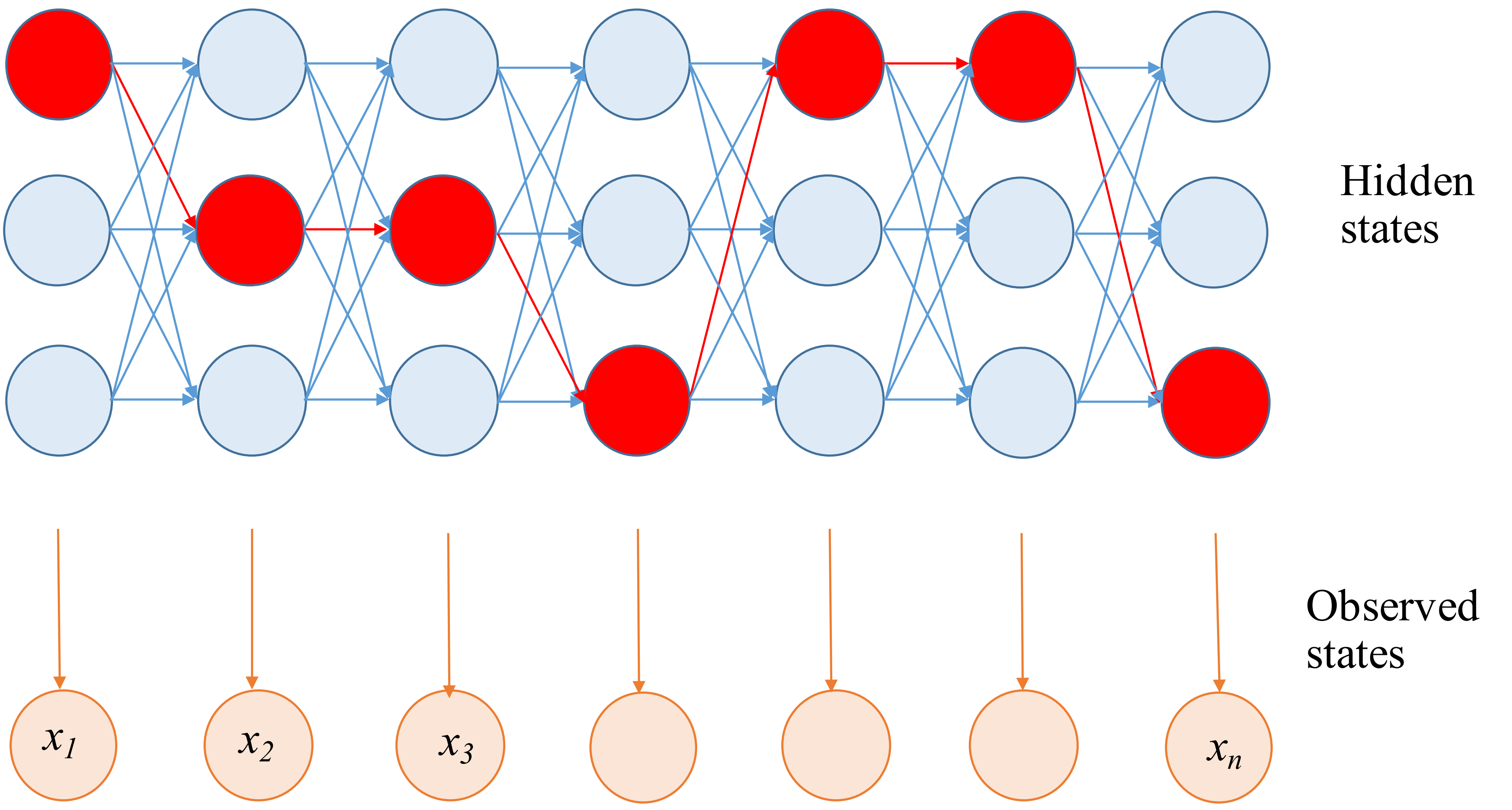
Developed by Andrew Viterbi, 1967

Andrew Viterbi,
Electrical Engineering
Department, USC

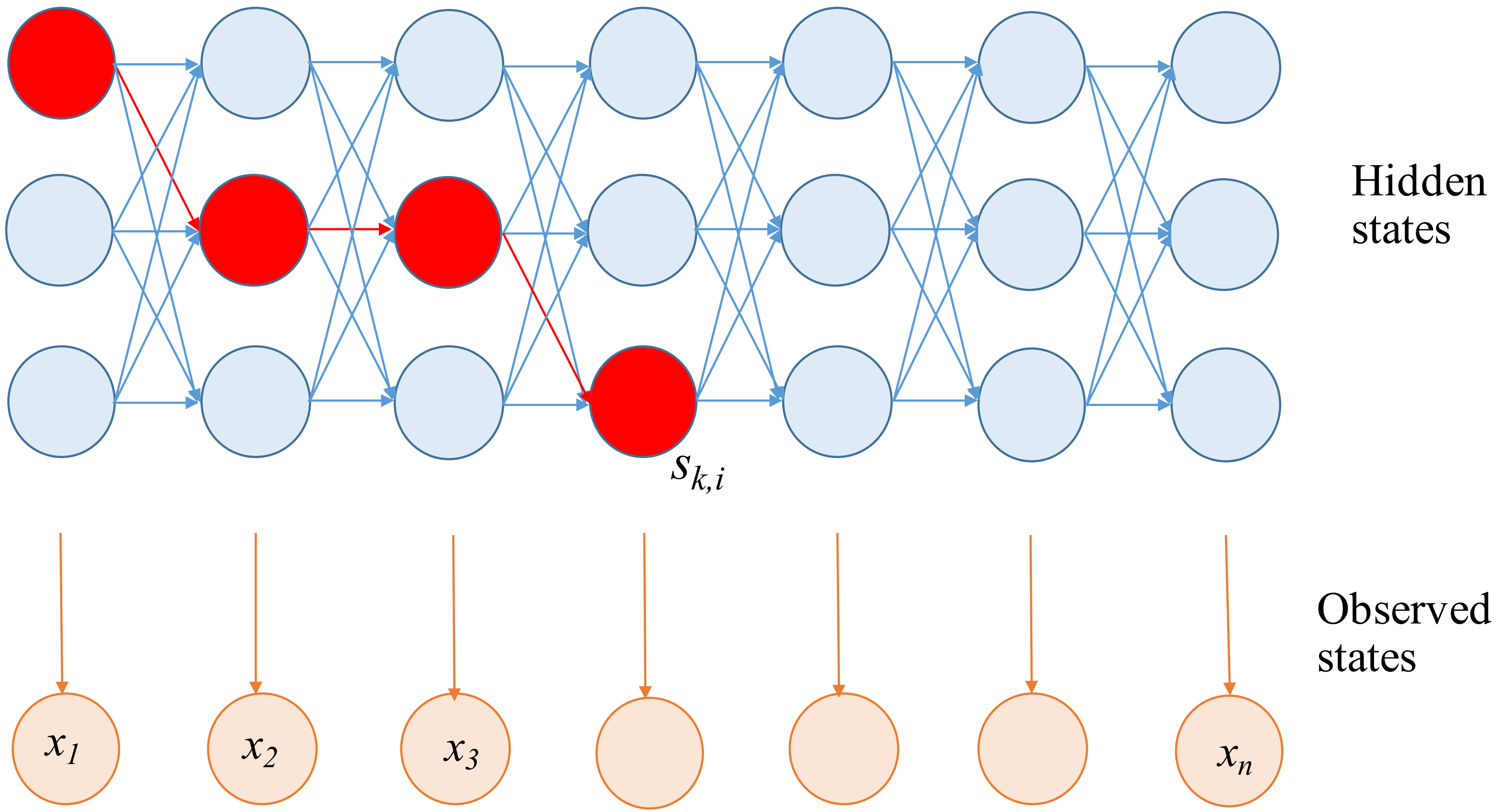


# Viterbi Decoding

Given the observed states, what is the most likely path in the hidden state ?

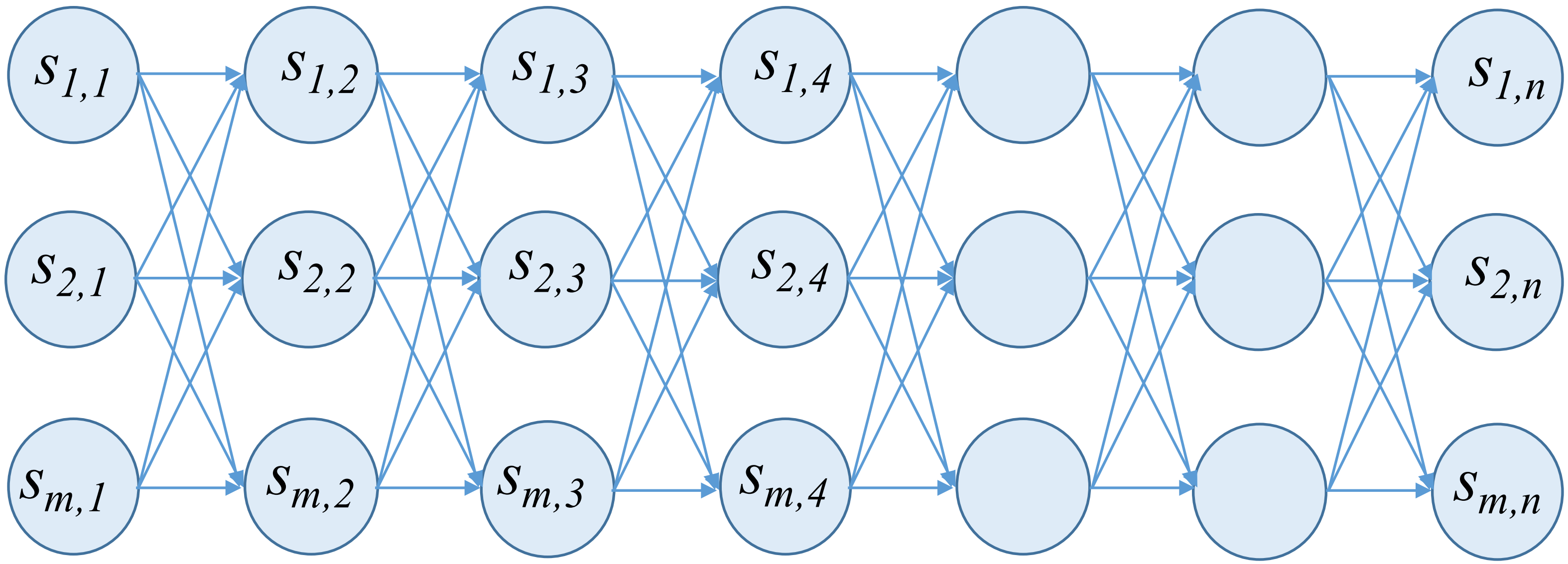


# Viterbi Decoding



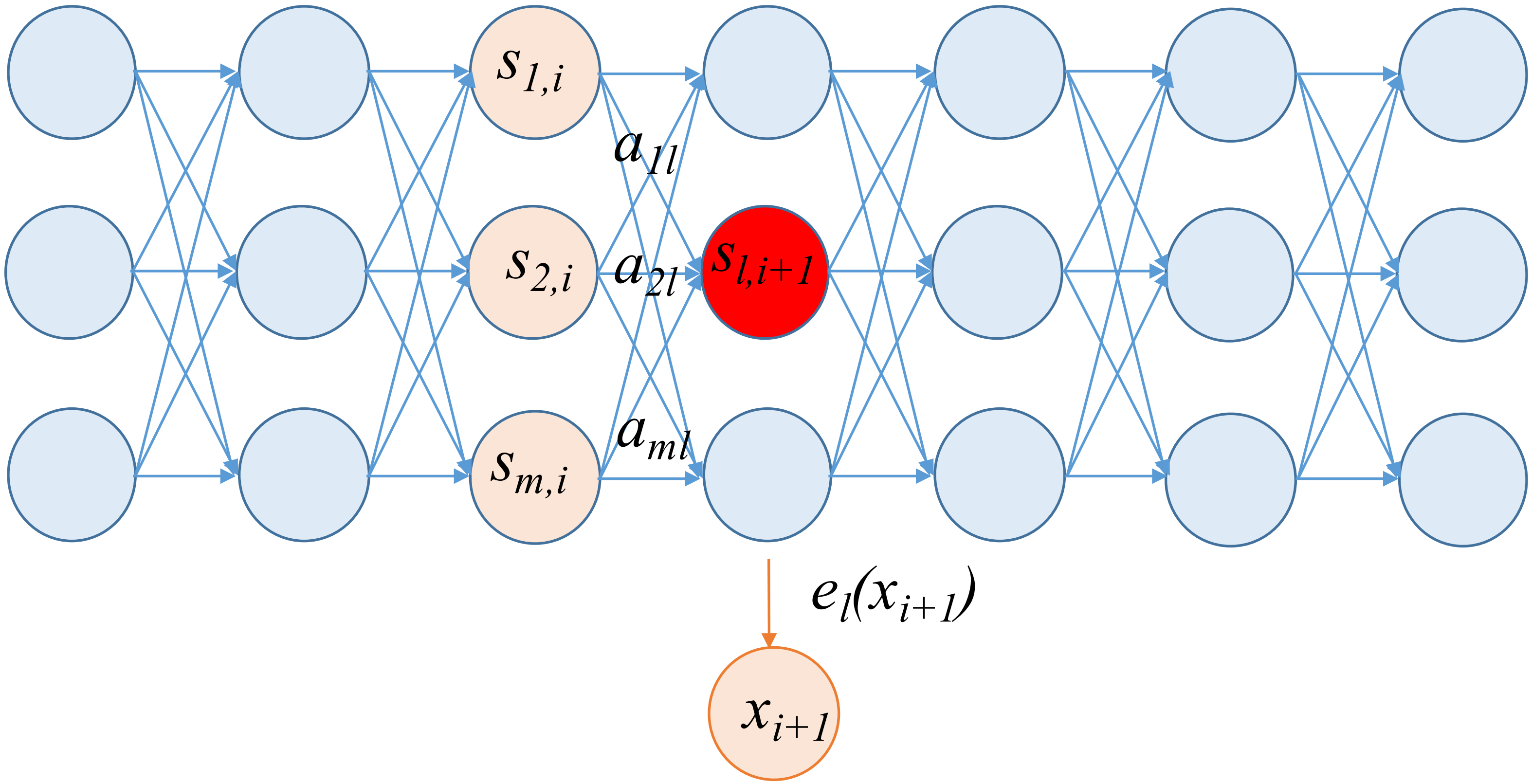

$s_{k,i}$ probability of the most likely path to state k$\in Q$ at step $1 \leq i \leq n$, given the observation $x_1\ x_2 \dots x_i$

# Viterbi Decoding




$s_{k,i}$ probability of the most likely path to state k$\in Q$ at step $1 \leq i \leq n$, given the observation $x_1\ x_2 \ldots x_i$

# Viterbi Decoding

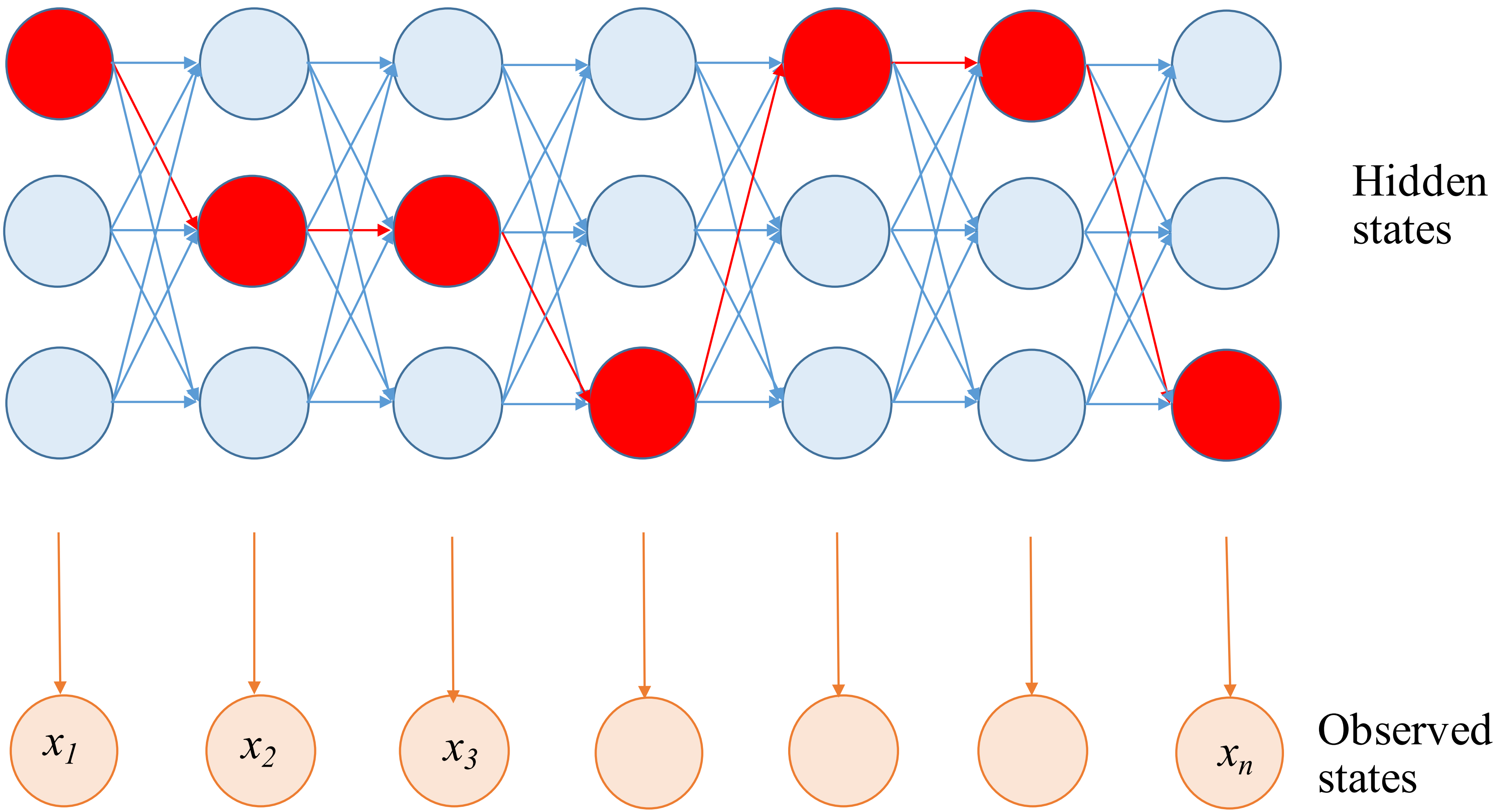



The most likely path to $l,i+1$ passes from $k,i$ for some $1 \leq k \leq m$

The likelihood of the optimal path to $l,i+1$ equals the maximum of the likelihood of the path $k,i$ , times the transition probability from state k to state l, and the emission probability $e_l(x_{i+1})$ for $1 \leq k \leq m$

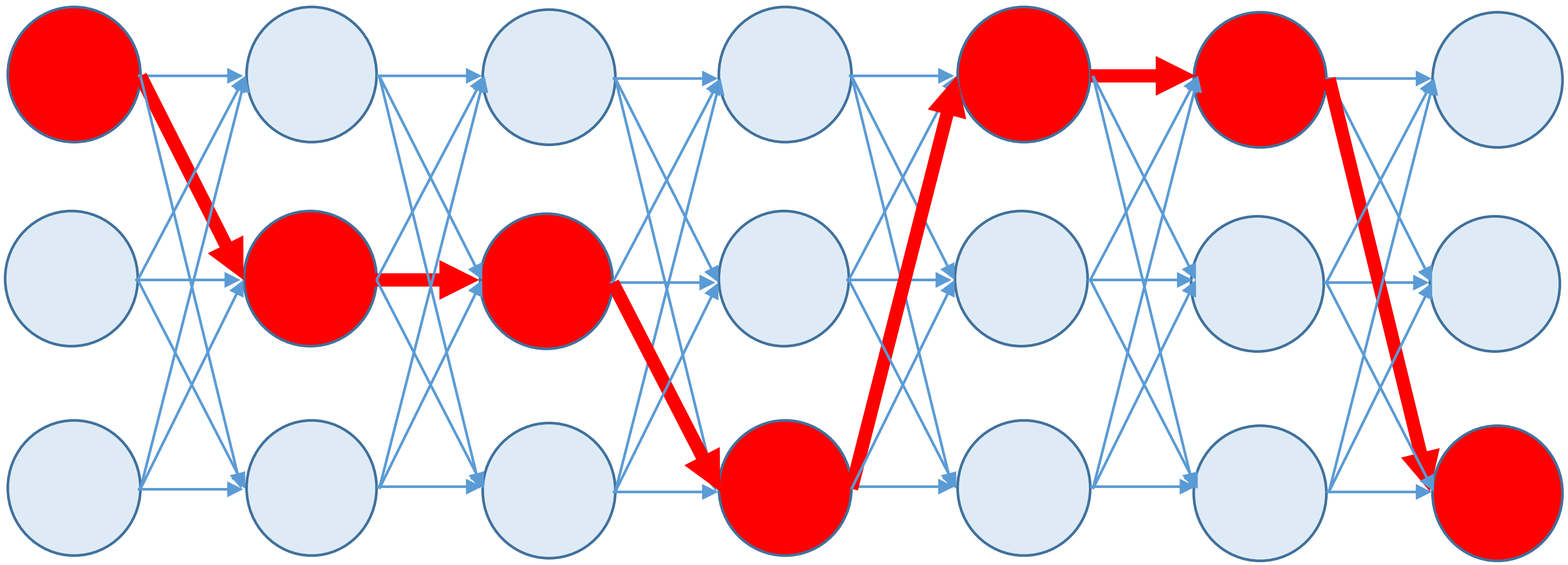
$$s_{l,i+1} = \max_{k \epsilon Q}\{s_{k,i} . a_{kl} . e_l\,(x_{i+1})\}$$

# Viterbi Decoding




probability of the optimal path at last step $n$, is the maximum $s_{k,n}$ for $k \in Q$

# Finding the optimal path



By recording the maximal state at each step, we can compute the hidden states $\pi_1\pi_2 \ldots \pi_n$ that maximize $p(x_1\, x_2 \ldots x_n|\pi_1\pi_2 \ldots \pi_n)$

# Forward & Backward algorithms

- Given observation $x$ and stat $k$, what is the probability $P(\pi_i=k|x)$ that the HMM was at state $k$ at time $i$ ?

$$P(x|\pi_i=k) = P(x_1,\ldots,xn|\pi_i=k) = P(x_1,\ldots,xi|\pi_i=k)\, P(x_{i+1,\ldots,}xn|\pi_i=k) = f_k(i)b_k(i)$$

- $f_{k,i}$ the probability of emitting prefix $x_1,\ldots,x_i$ given $\pi_i=k$.

- $P(\pi_i=k|x)$ is affected not only by the values $x_1,\ldots,x_i$ but also $x_{i+1},\ldots,x_n$

- $b_{k,i}$ the probability of emitting prefix $x_{i+1},\ldots,x_n$ given $\pi_i=k$.

- $f_{k,i}$ and $b_{k,i}$ using an iterative method very similar to what we saw in the previous slides.

# Forward & Backward algorithm

- $f_{k,i}$ can be computed using the following iterations :

$$f_{k,i} = \sum_{l \in Q} f_{l,i-1} . alk \ .ek \ (xi)$$

Maximization replaced by summation

- $b_{k,i}$ can be computed using the following iterations :

$$b_{k,i} = \sum_{l \in Q} b_{l,i+1} \ .akl \ .el \ (xi_{+1})$$

# Complexity

Viterbi decoding works in $n|Q|^2$ time

Because each step involves multiplication, we might end up with overflows (very small/large numbers)

To avoid overflows, we can switch to logarithmic space by defining $S_{k,i}=log\, s_{k,i}$

$$S_{l,i+1} = \max_{k \epsilon Q}\{S_{k,I} + \log(a_{kl}) + \log e_l\,(x_{i+1})\}$$

# Estimating HMM parameters

- How to estimate parameters from transition probabilities Q and and emission probabilities Σ ?

$x^1, \ldots, x^m$ : training sequence

Find maximum likelihood Σ and Q from $x^1, \ldots, x^m$

$max_{\Sigma,Q} \prod_{j=1}^{m} P(x^j \mid \Sigma, Q)$

Optimization in multi-dimensional parameter space $\Sigma, Q$

# Supervised & unsupervised estimation

- If for each $x^i$ , we now the corresponding hidden state $\pi^i$ , the problem is easy

- If $\pi^i$ are unknown, the problem is more difficult.

# Parameter estimation in Supervised case

- $A_{kl}$ : the number of transitions from state $k$ to state $l$
- $E_k(b)$ : the number of times $b$ emitted from state $k$

$$a_{kl} = \frac{A_{kl}}{\sum_{q \epsilon Q} A_{kq}} \qquad e_k(b) = \frac{E_k(b)}{\sum_{\sigma \epsilon \Sigma} E_k(\sigma)}$$

# Baum-Welch iterations (unsupervised case)

- Guess initial state labels $\pi^i$ for each input $x^i$ and perform the following iterations :

  - Estimate $\Theta$ from labels $\pi^i$

  - Compute optimal states $\pi^i$ from $\Theta$ and $x^i$ using Viterbi decoding