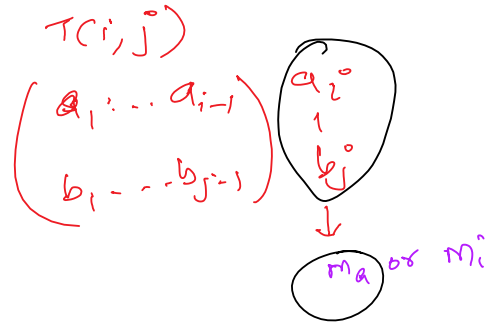
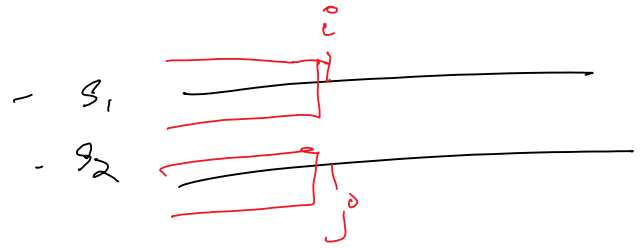
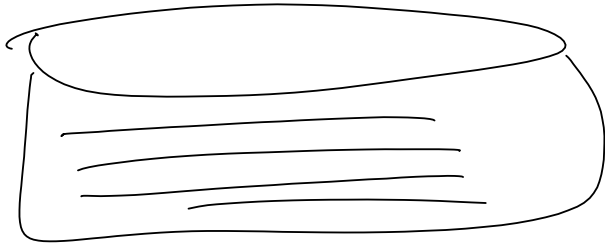


Probabilistic Modeling

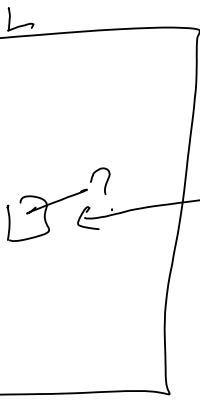
Friday, April 3, 2020 10:39 AM



bacterial genomes

A A A A
 | | | |
 A C G T

20



M
 |
 L

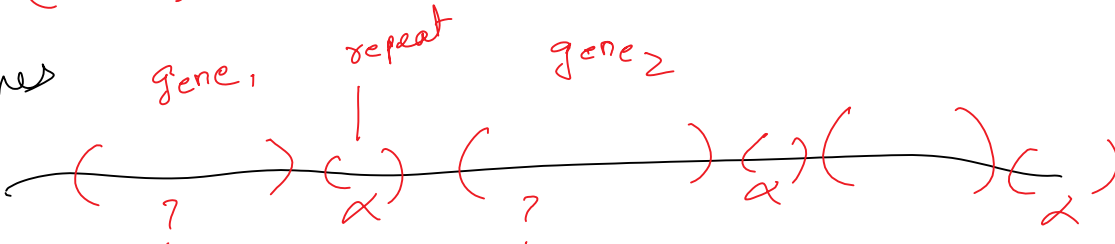
20

M

Speech recognition



Genomes



Observation:

$$x = x_1, x_2, x_3, \dots, x_i, \dots, x_L$$

Likelihood ratios

Friday, April 3, 2020 10:54 AM

Basic Probability:

$p(x)$ ← prob. of event x
"observation"

$p(x, y)$ ← prob. of x & y
(Joint prob.)

$$\begin{aligned}
 p(x, y) &= p(x|y) p(y) \\
 &= p(y|x) p(x) \\
 \Rightarrow \text{Bayes Theorem} \\
 p(x|y) &= \frac{p(y|x) p(x)}{p(y)}
 \end{aligned}$$

conditional prob.

Let Model M (be "our" model) ← alternative hypothesis
 Model R ← be random model ← null hypothesis

$p(x|M)$ ← prob. of observing x under M

$p(M|x)$ ← prob. of Model M given x (posterior prob.)

(prior probability)

Likelihood Ratios: / odds

$$\begin{aligned}
 \text{L.R.} : \frac{p(x|M)}{p(x|R)} &= \frac{p(M|x) p(x)}{p(R|x) p(x)} \cdot \frac{p(R)}{p(M)} \\
 &= \frac{p(M|x)}{p(R|x)} \cdot \frac{p(R)}{p(M)}
 \end{aligned}$$

(Ratio of the posteriors)

Log Likelihood Ratios (LLR)

assumption

if $p(R) = p(M)$

$$\Rightarrow \frac{p(x|M)}{p(x|R)} = \frac{p(M|x)}{p(R|x)}$$

$$\begin{aligned}
 p(x|M) &= p(x_1, x_2, \dots, x_L | M) \\
 &= p(x_1|M) p(x_2|x_1) p(x_3|x_1, x_2) \dots \\
 &\quad \dots p(x_L|x_1, \dots, x_{L-1})
 \end{aligned}$$

$$= p(x_1|M) \prod_{i=2}^L p(x_i | x_1, \dots, x_{i-1})$$

$$\begin{aligned}
 \log(p(x|M)) &= \log(p_1|M) + \sum_{i=2}^L \log(p \dots)
 \end{aligned}$$

$p(x|R)$

$$\log AB = \log A + \log B$$

Discrimination

Friday, April 3, 2020 10:54 AM

previous prob.

→ Casino example

☰ → loaded die

☲ → fair die

$$p(6) = \frac{1}{2}, p(1) = \dots = p(5) = \frac{1}{10}$$

$$p(1) = p(2) = \dots = p(6) = \frac{1}{6}$$

$x \in \{L, F\}$
 $x = x_1, x_2, x_3, \dots, x_i, \dots, x_L$

$$\frac{p(x|L)}{p(x|F)} = \frac{p(L|x)}{p(F|x)}$$

// assuming $p(F) = p(L)$
 $p(x_i|L) = \frac{1}{2}$ (if $x_i = 6$)

Example:

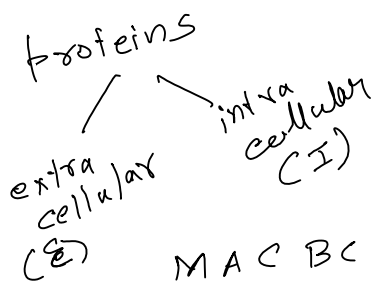
Observation: $x = 3 \overset{x_1}{1} \overset{x_2}{2} 6 6 2 6 3 4 2 \overset{x_p}{6} 6 6 6 2 6 1 6 4$
 $L \leftarrow \# \text{ len. obs.}$

$$p(x|L) = p(x_1|L) \prod_{i=2}^L p(x_i | x_1, \dots, x_{i-1}, L)$$

$$p(x|F) = p(x_1|F) \prod_{i=2}^L p(x_i | x_1, \dots, x_{i-1}, F)$$

$$\text{Likelihood Ratio} = \frac{p(x|L)}{p(x|F)}$$

Bio. Example:



amino acid $\Sigma = \{20 \text{ aa}\}$
 ↓
 M A B B B B P L M L P . . .

M A C B C V B P L M . . .

$$\frac{p(x|E)}{p(x|I)}$$

Markov Chains

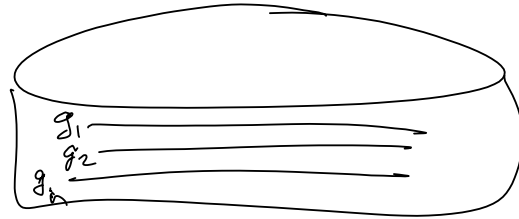
Friday, April 3, 2020 10:54 AM

Gene Identification:

x : A C C C G C G A C G T A T C G T

x_L

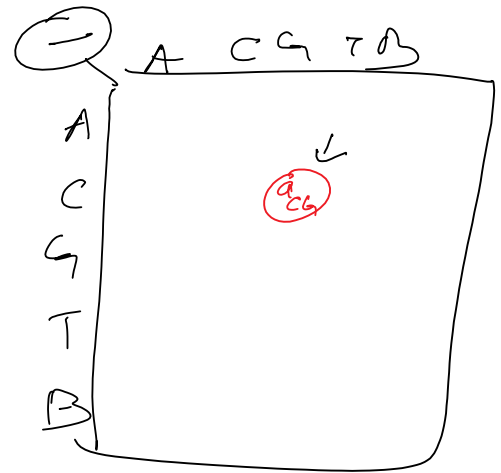
Is $x \in \text{Gene}$? $\rightarrow M$
 or (not)? $\rightarrow R$



database

Let $(+)$ \leftarrow model of all genes

$(-)$ \leftarrow not a gene



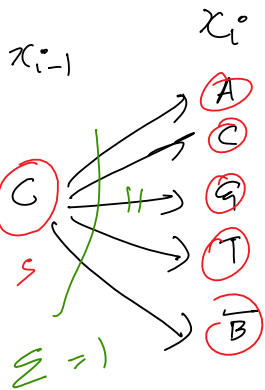
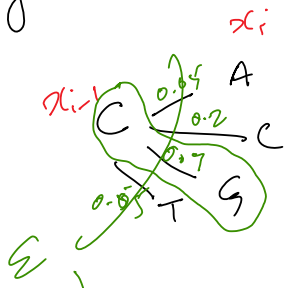
transition table:

x_i

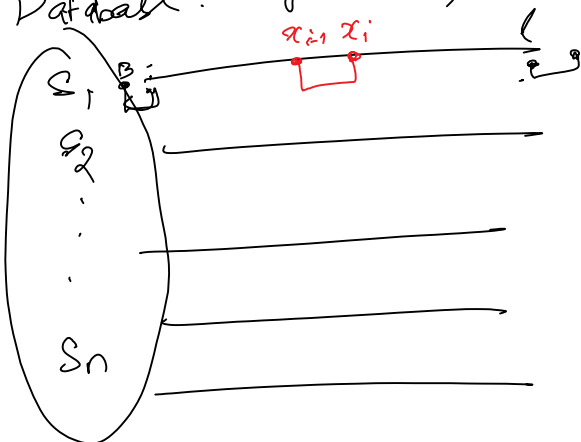
	A	C	G	T	B
A					
C	a_{CA}	a_{CC}	a_{CG}	a_{CT}	a_{CB}
G					
T					
B					

$= 1$

Let a_{sc} \leftarrow prob. of transition from "state" s to state t



Database: All genes (+)



'c's are followed by 's'

$$P(CG | +) = \frac{a_{CG}}{\left(\sum_{i=1}^n |S_{g_i}| + n \right)}$$

\nearrow
 a_{CG}

Markov Chains

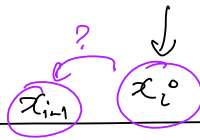
Friday, April 3, 2020 10:54 AM

$$x = \overset{0}{B} \overset{\downarrow}{x_1} \overset{\cdot}{x_2} \dots \overset{\cdot}{x_L} \overset{0}{B}$$

$$p(x|+) = a_{B x_1}^+ \prod_{i=2}^{L+1} a_{x_{i-1} x_i}^+$$

$$p(x|-) = a_{B x_1}^- \prod_{i=2}^{L+1} a_{x_{i-1} x_i}^-$$

Markov Chains -



- $x_i \leftarrow$ depends on $x_{i-1} \Rightarrow$ 1st order Markov chain
- $x_i \leftarrow$ depends on $x_{i-1} x_{i-2} \Rightarrow$ 2nd order MC
- $x_i \leftarrow$ depends on k prev. symbols $\Rightarrow k^{\text{th}}$ order MC

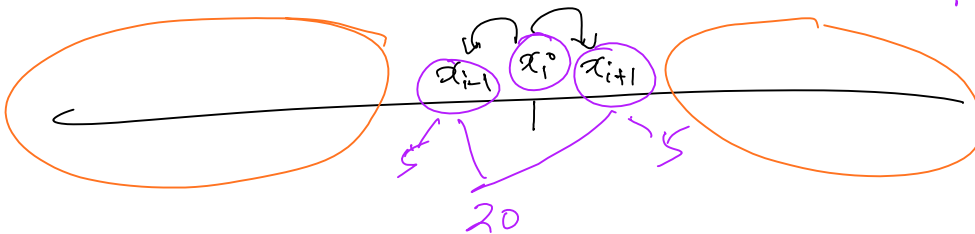
$x_i \leftarrow$ depends on all prev. observations

$$p(x_i | x_1 \dots x_{i-1}) = p(x_i | x_{i-1})$$

↓ don't care

context

$$x_i = f(x_{i-1} x_{i+1})$$



Markov Chains

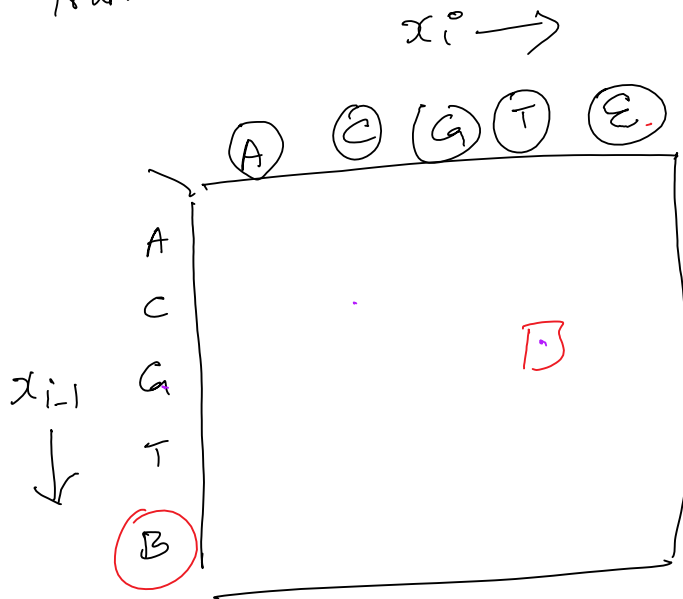
Friday, April 3, 2020 10:54 AM

Markov Chains (Definition):

- Σ : symbols
- \mathbb{I} : set of states
- Q_{st} : transition probability from state s to state t

CpG Islands Example:
(Gene Identification)

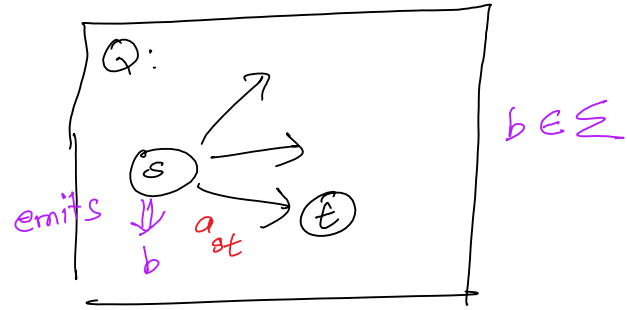
Transition Table:



$$\Sigma = \{A, C, G, T, B, \epsilon\}$$

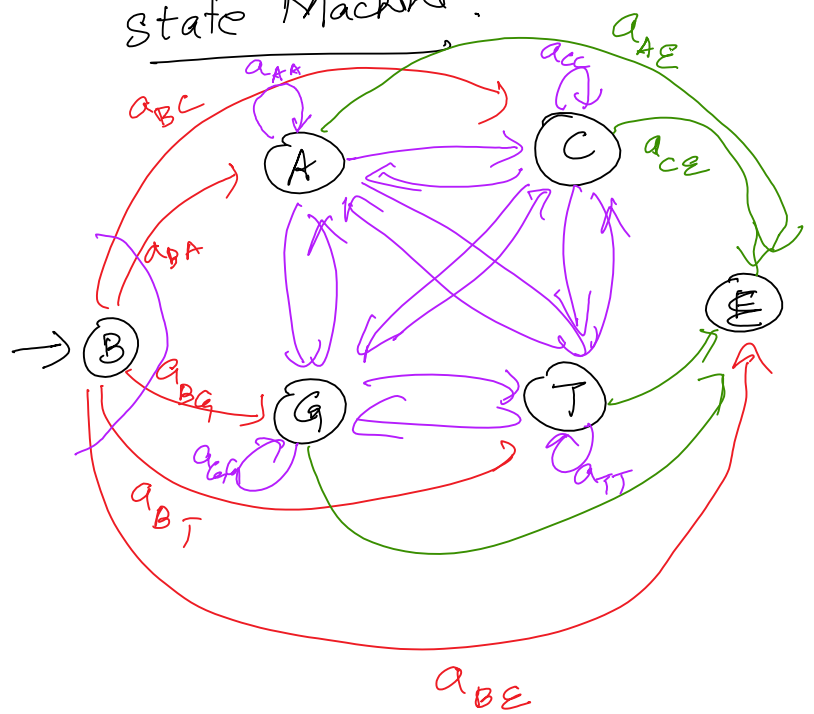
$$\mathbb{I} = \{A, C, G, T, B, \epsilon\}$$

Abstraction as a state machine:



Key Markovian assumption:
 \equiv symbol at i^{th} state depends on the symbol at the $(i-1)^{\text{th}}$ state

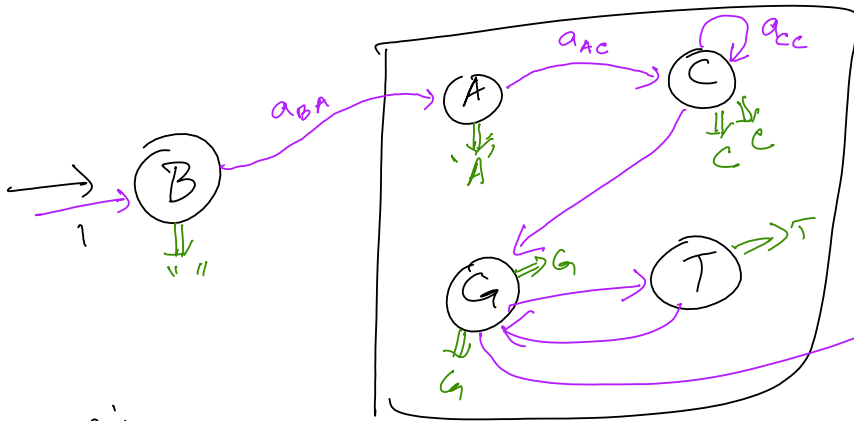
state Machine:



Emitting symbols for states

Friday, April 3, 2020 10:54 AM

Let $e_k(b) \leftarrow$ prob. of emitting b from state k



$e_A(A) = 1$
 $e_{A(C)} = 0$
 $e_A(G) = 0$
 $e_{\#(T)} = 0$

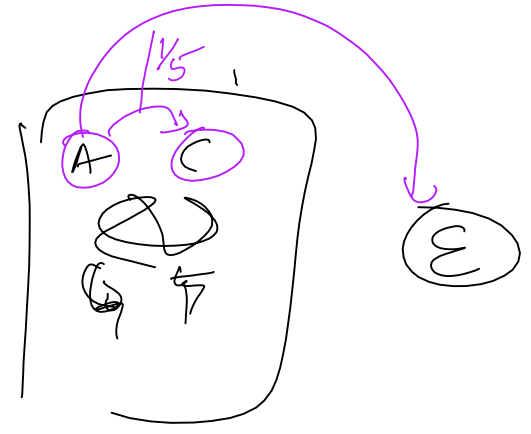
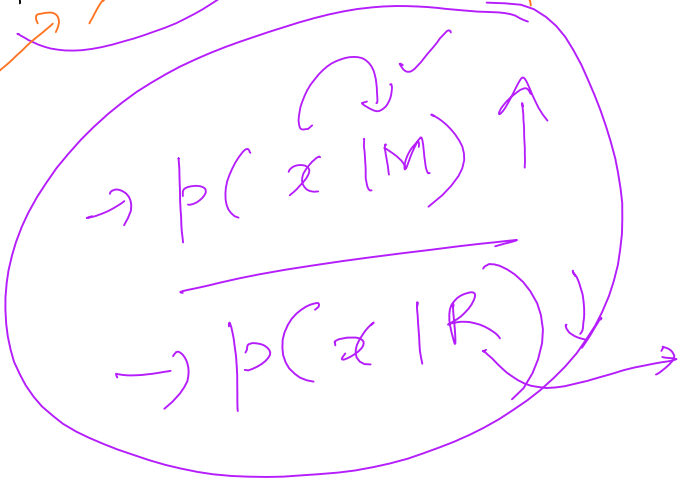
Observation:

$x = \text{A C C G T G}$

- > Every state can emit one symbol at step i
- > emitted symbol = observed symbol at x_i

$p(x, \pi)$

$p(x | M) = 1 \cdot a_{BA} \cdot a_{AC} \cdot a_{CC} \cdot \dots \cdot a_{GT} \cdot a_{TG} \cdot a_{GE}$



Is $x \in M$?
 or $x \in R$?

discrimination

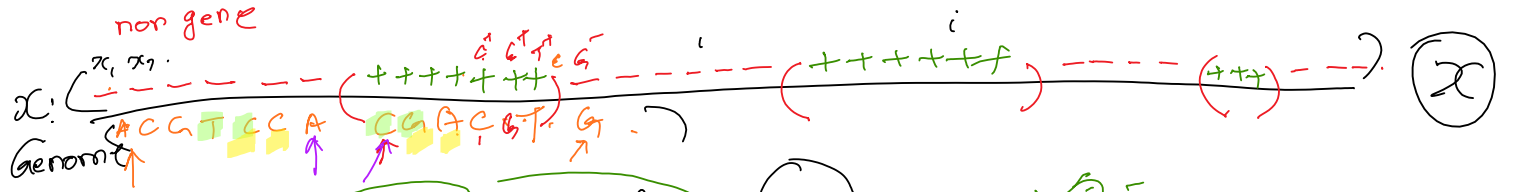
HMMs: Motivation

Friday, April 3, 2020 10:54 AM

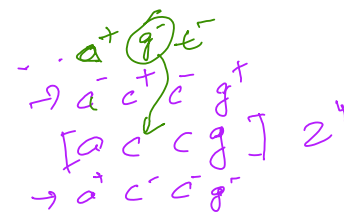
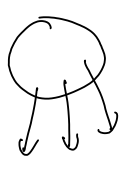
Motivation:

"Decoding" a sequence of observations

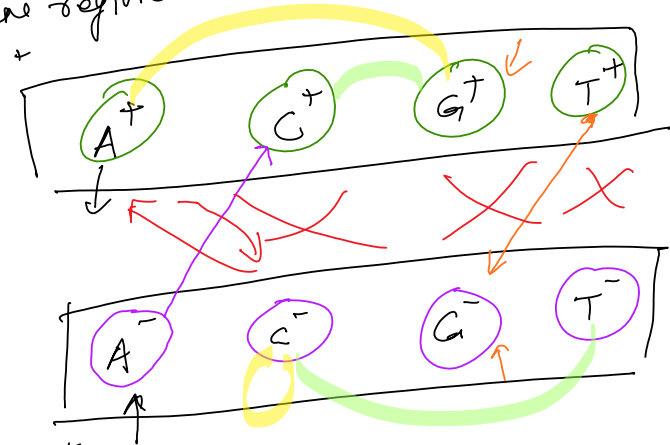
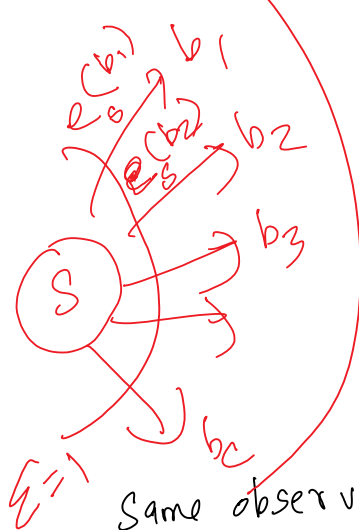
Q) "decode" the genetic regions.



decouple state \neq symbol



Let +: gene region
-: non gene region

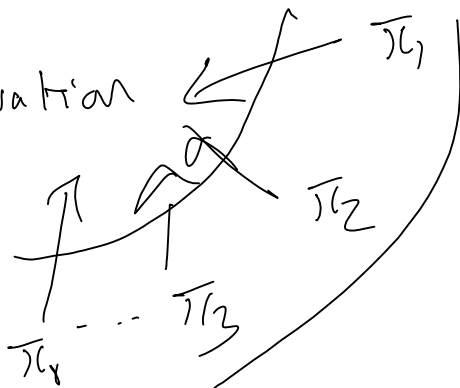


$$\Sigma = \{b_1, b_2, \dots, b_C\}$$

$$Q = \{1, 2, \dots, K\}$$

$$K \neq C$$

Same observation (x)



params

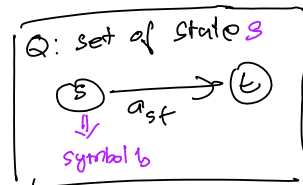
$$\left\{ \begin{array}{l} p(x, \pi_1) \\ p(x, \pi_2) \\ p(x, \pi_3) \\ \vdots \\ p(x, \pi_r) \end{array} \right\}$$

max path π^*

$\pi_{2nd\ best}$

HMMs: Overview

Friday, April 3, 2020 10:54 AM



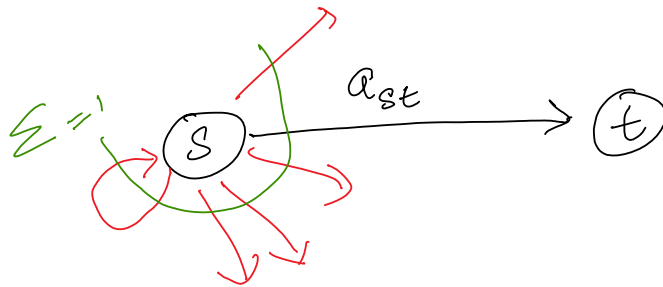
HMM Main Features:

A finite state machine with transition and emission probabilities.

1. Transition defined from every state (S) to every other state (t) with 0 or more probability.

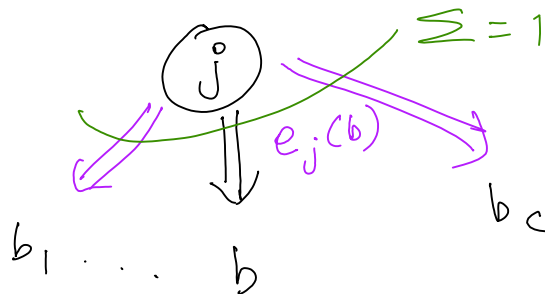
s.t.

$$\sum_{t \in Q} a_{st} = 1$$



2. Any state (j) can emit any symbol (b), with a predefined probability ($e_j(b)$)

s.t. $\sum_{b \in \Sigma} e_j(b) = 1$, for any state j



Hidden Markov Models (HMMs)

Friday, April 3, 2020 10:54 AM

A HMM Model should contain:

components

Σ : $\{b_1, b_2, \dots, b_c\}$: the alphabet of c symbols

Q : $\{1, 2, \dots, k\}$: a set of k states

weights

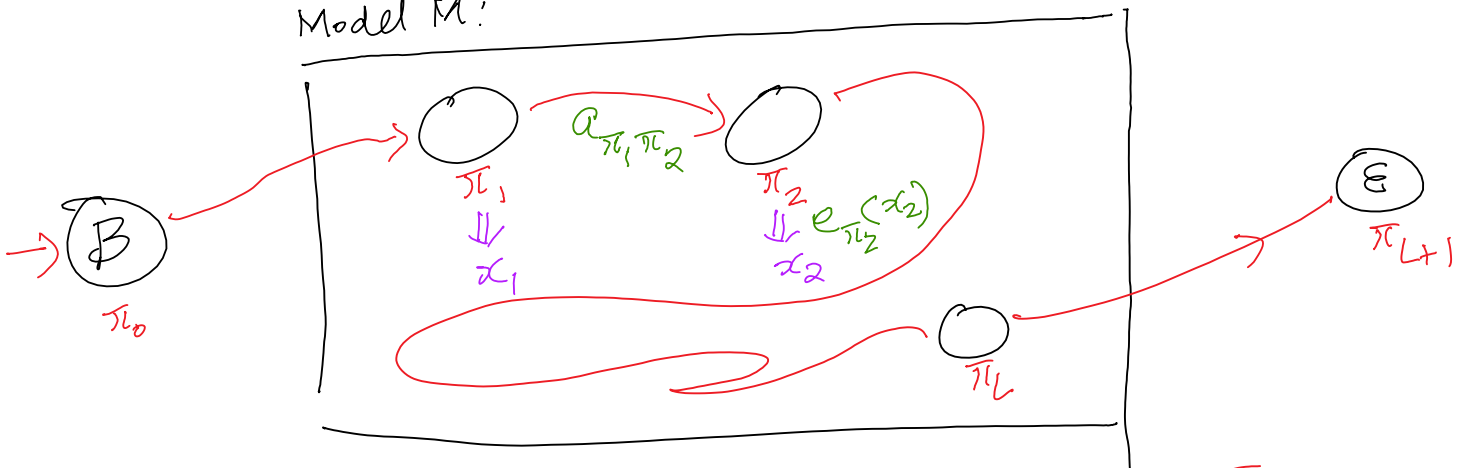
a_{st} : transition probability from state s to state t

$e_j(b)$: emission probability of symbol b from state j

Model Operation:

When presented with a sequence x of observations ($x = x_1, x_2, \dots, x_L$),
 The model M generates x by following a path π (of length $L+1$)
 ($\pi = \pi_0 \pi_1 \pi_2 \dots \pi_L \pi_{L+1}$)

Model M:



π : $\pi_0 \pi_1 \pi_2 \dots \pi_i \dots \pi_L \pi_{L+1}$
 x : $- [x_1 \ x_2 \dots x_i \dots x_L] -$

$p(x, \pi) \leftarrow$ joint prob of generating x
 by traversing path π in M .

Viterbi's decoding

Wednesday, April 8, 2020 10:48 AM

Viterbi's decoding (main ideas):

Input:

Model $M(\text{HMM}) : \{ \Sigma, Q, \delta, \epsilon \}$: transition prob; ϵ : emission prob.

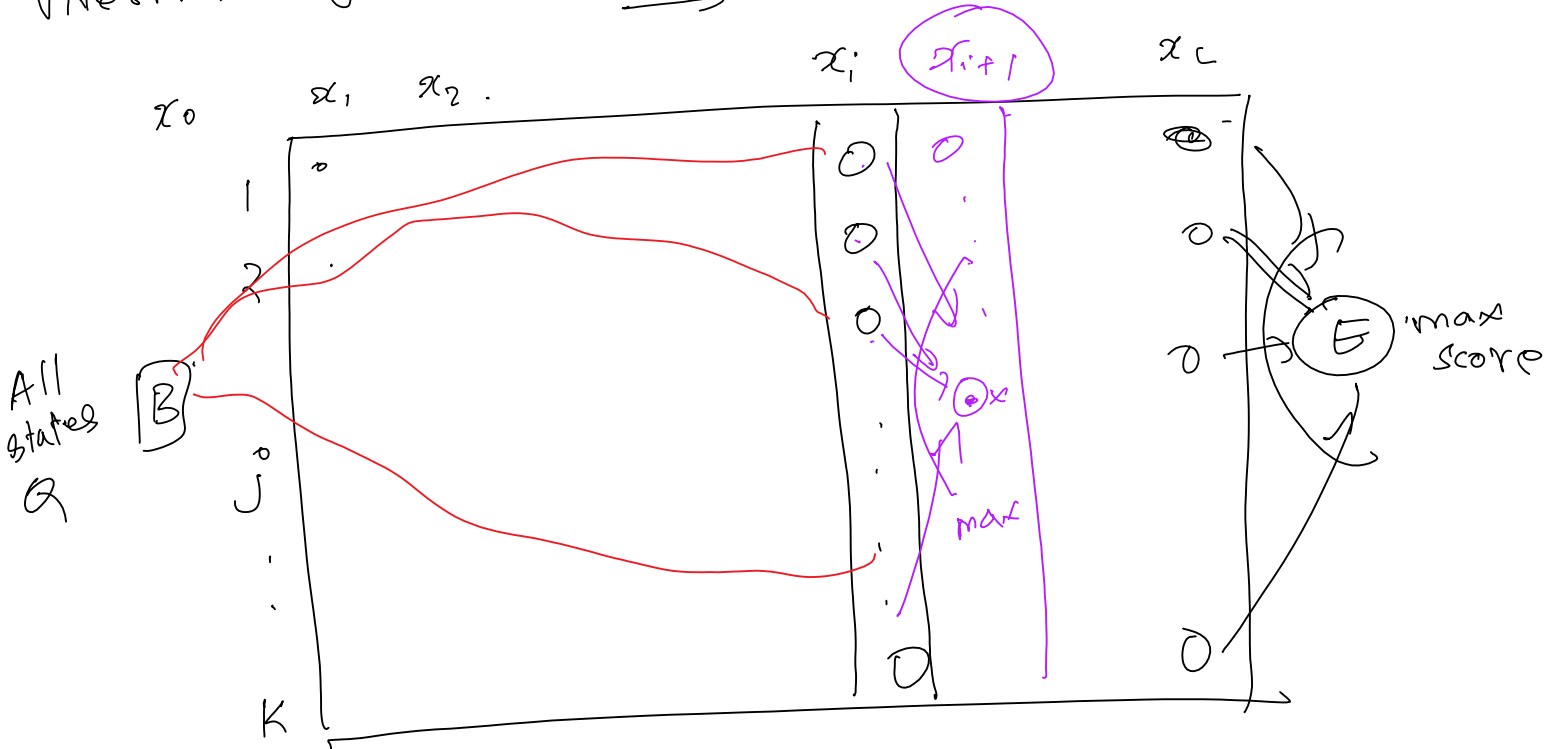
Observation x : $x_1 \quad \dots \quad x_i \quad \dots \quad x_L$

Output path π^* : $B \quad \pi_1 \quad \pi_2 \quad \pi_3 \quad \dots \quad \pi_i \quad \dots \quad \pi_L \quad E$

(max prob)

Viterbi's Algo:

time steps
→



Viterbi's Decoding

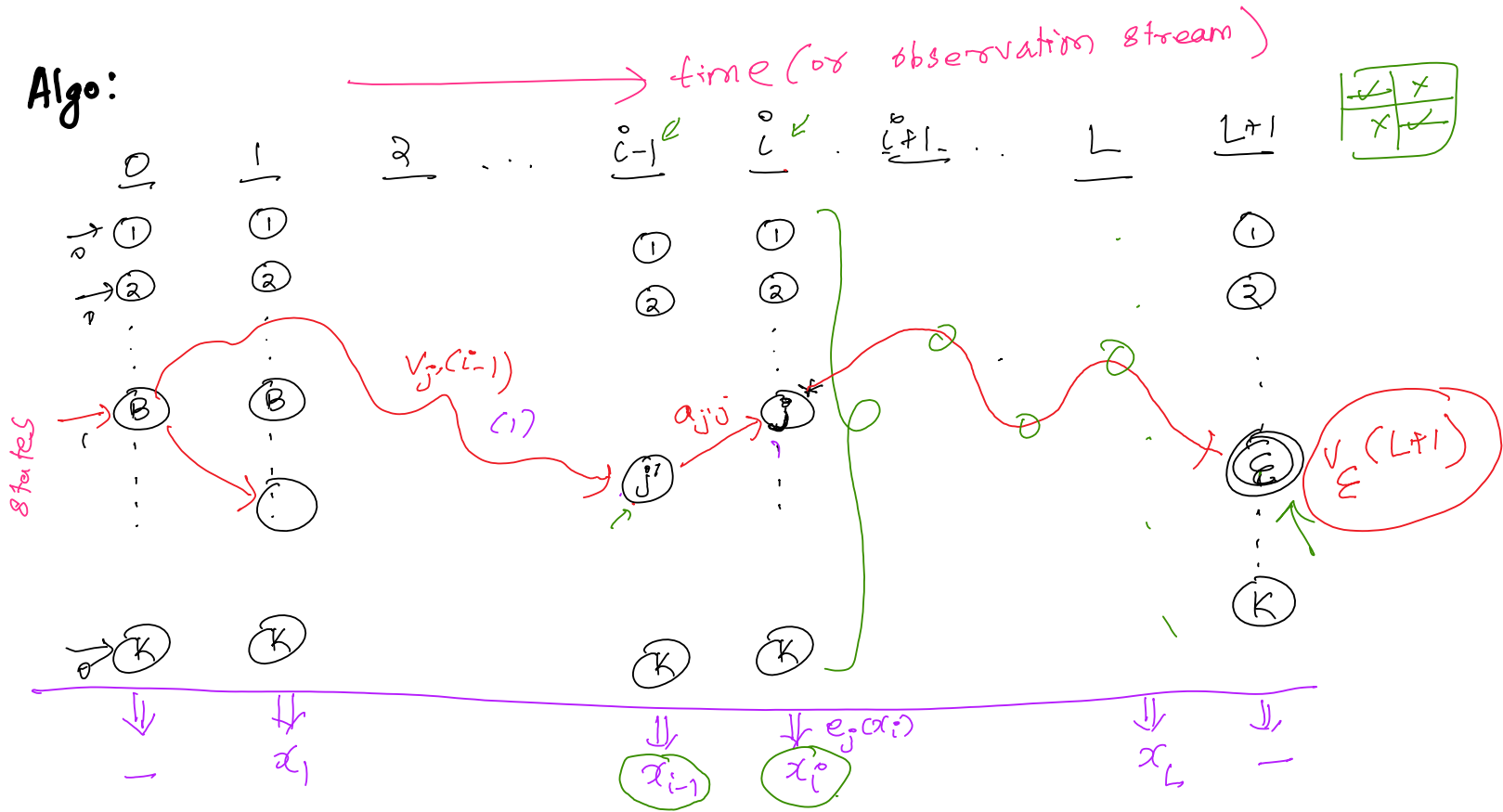
Friday, April 3, 2020 10:54 AM

Input: Model M , observation $x = x_1, \dots, x_L$

Output: $\pi^* \leftarrow$ path that maximizes the joint probability $p(x, \pi)$

Goal: $\pi^* \leftarrow \arg \max_{\pi} p(x, \pi)$

Algo:



Let $v_j(i) \leftarrow$ max prob. for generating (x_1, x_2, \dots, x_i) using a path that ends at state j .

1) **Init:** $v_B(0) = 1$

for $(i = 1 \text{ to } L+1)$

$O(L \times K \times K) = O(LK^2)$: Time complexity
 $O(LK)$: Space complexity

2) **Recurrence:**

for $(j = 1 \text{ to } K)$

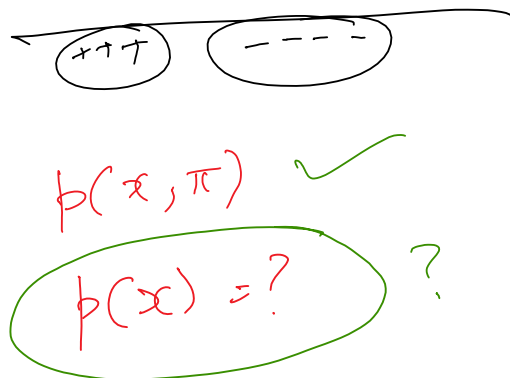
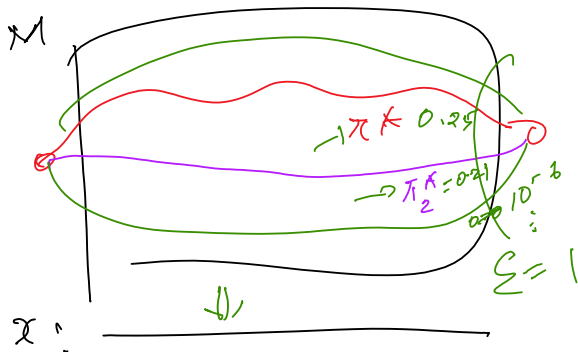
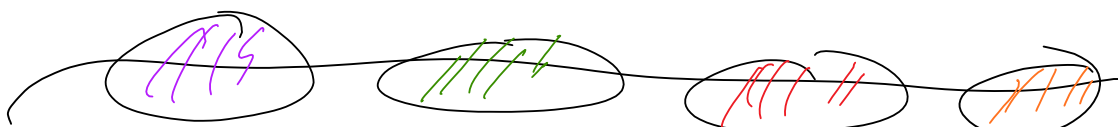
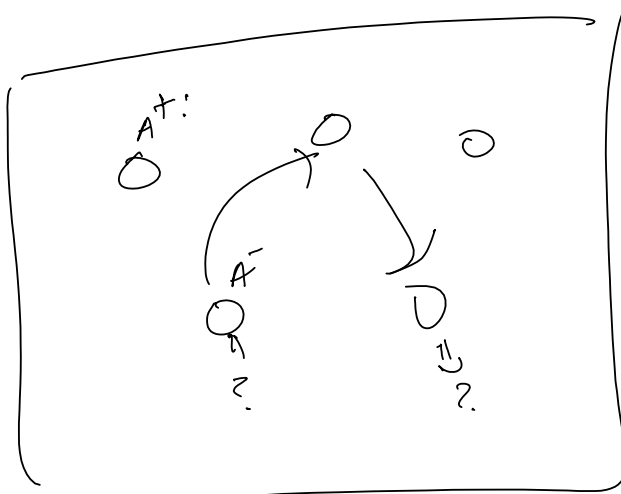
$$v_j(i) = \max_{1 \leq j' \leq K} \{ v_{j'}(i-1) \cdot a_{j,j'} \cdot e_j(x_i) \}$$

Output $v_E(L+1) \Rightarrow$ max probability path (π^*) score

3) **Retrace** from $v_E(L+1)$ back to $v_B(0)$ to reconstruct the opt. path π^* (decoded answer)

Decoding and other functions

Wednesday, April 8, 2020 10:48 AM



$\phi(x) = 0.2$

$p(x) = \text{Sum} \left\{ \begin{array}{l} p(x, \pi_1) \\ p(x, \pi_2) \\ \vdots \\ p(x, \pi_{exp}) \end{array} \right\}$

\downarrow

$p(x|M)$

Forward Algorithm

Friday, April 10, 2020 9:46 AM

Input: Model M ; Observation $x = x_1 \dots x_L$

Output: $p(x|M)$

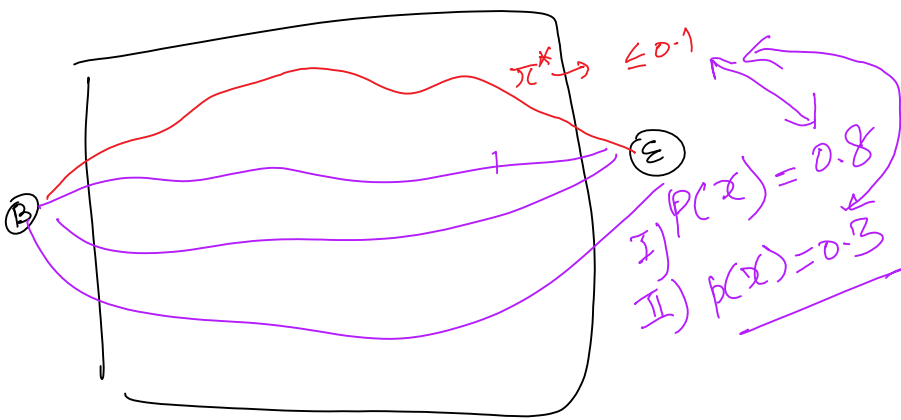
// over all possible paths that you can take on model M to generate x .

Motivation:

→ π^* from Viterbi's decoding gives only the maximum.. probability path.

a) What if that probability is itself not so "high"?

b) We need a way to compute $p(x)$ for HMMs (similar to how we computed $p(x)$ for Markov chains.



For Markov chains:

$$p(x) = p(x_1, x_2, x_3, \dots, x_L) \\ = p(x_1) p(x_2|x_1) \dots \\ p(x_L|x_{L-1})$$

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

$$p(x) = \text{sum} \left\{ \begin{array}{l} p(x, \bar{x}_1) \\ p(x, \bar{x}_2) \\ \vdots \\ p(x, \pi^*) \\ \vdots \\ p(x, \bar{x}_{end}) \end{array} \right\}$$

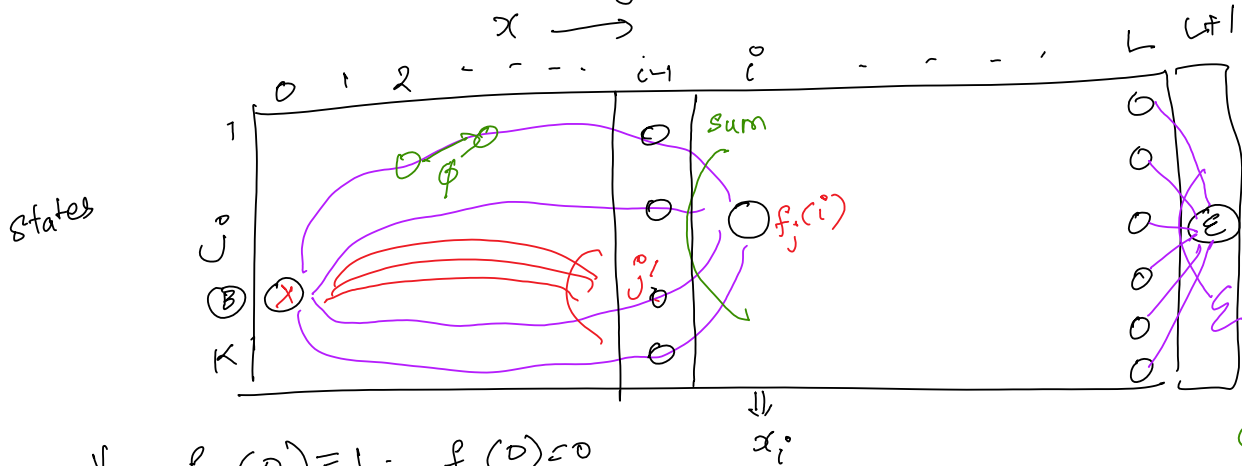
Forward Algo

Friday, April 10, 2020 9:46 AM

Forward Algo:

Input: Model M , observation x .
Output: $p(x|M)$

Let $f_j(i) \leftarrow$ probability of generating x_1, \dots, x_i using any path ending at state j .



Init: $f_B(0) = 1$; $f_j(0) = 0$ for $j \neq B$

for $j = 1$ to K

$$f_j(i) = \sum_{j'=1}^K f_{j'}(i-1) a_{j'j} e_j(x_i)$$

Complexity:
 $O(KL \times K)$
 $= O(LK^2)$

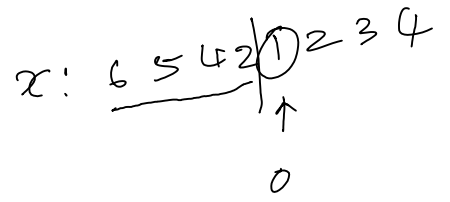
Only forward computing necessary

Output $f_E(L+1) \equiv p(x)$

$$L = p(6) = \frac{1}{2}$$

$$p(1) = 0$$

$$p(2|3|4|5) = \frac{1}{8}$$

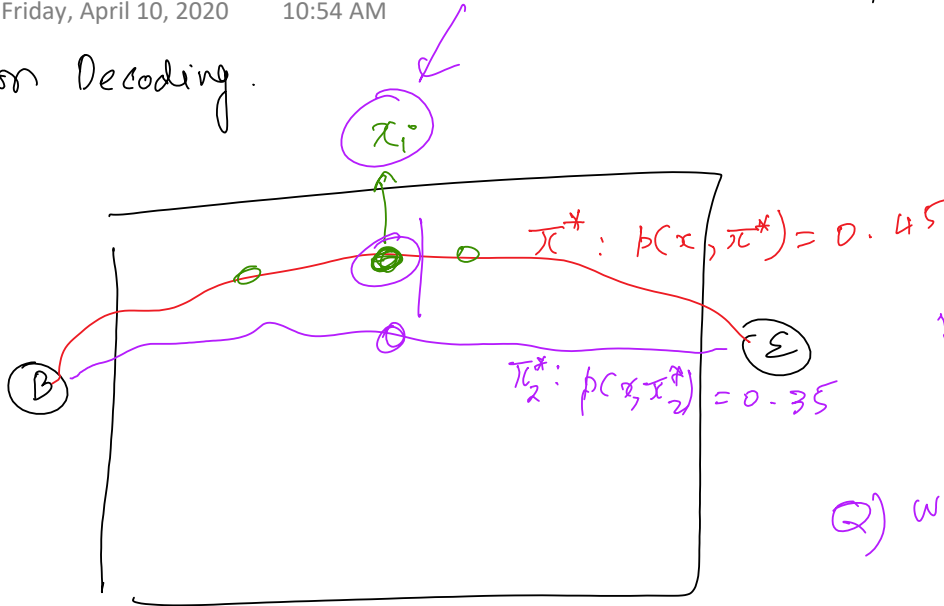


Posterior Decoding

Friday, April 10, 2020 10:54 AM

forward Alg:
 $p(x|m)$

Posterior Decoding.



$p(x, \pi)$
 ↑
 global x

Q) what is the best explanation for x_i ?

local specific question

$$p(\pi_i = j | x) \quad ?$$

$x: x_1, x_2, \dots$

x_i

$\pi_i = j$

$\text{Arg Max}_j p(\pi_i = j | x)$

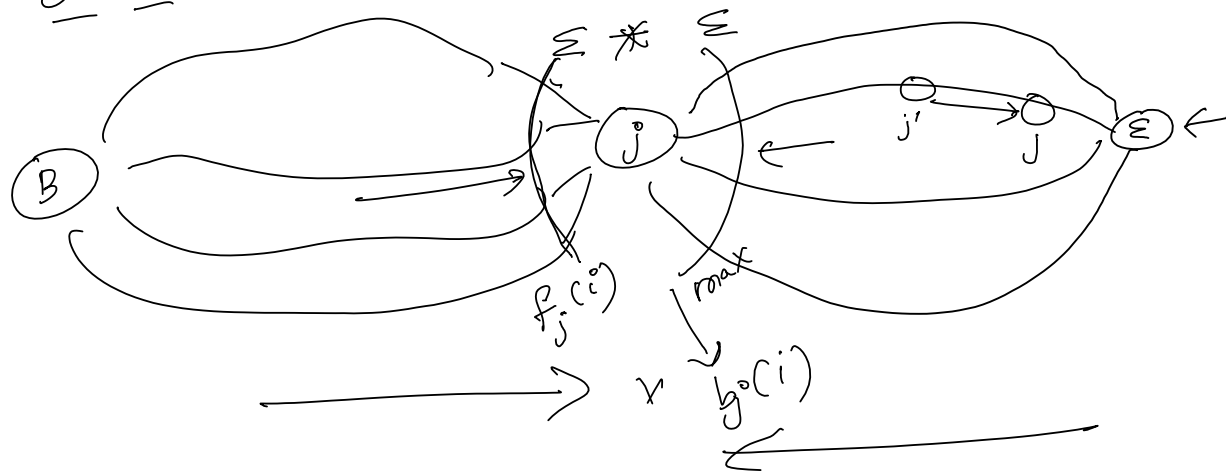
x_L

Best state to explain x_i =

Posterior decoding

Wednesday, April 15, 2020 10:57 AM

$$p(\pi_i = j | x) = p(x_1, \dots, x_i | \pi_i = j) \times \frac{p(x_{i+1}, \dots, x_L | \pi_{i+1} = j)}{p(x_{i+1}, \dots, x_L)}$$



Paired HMM

