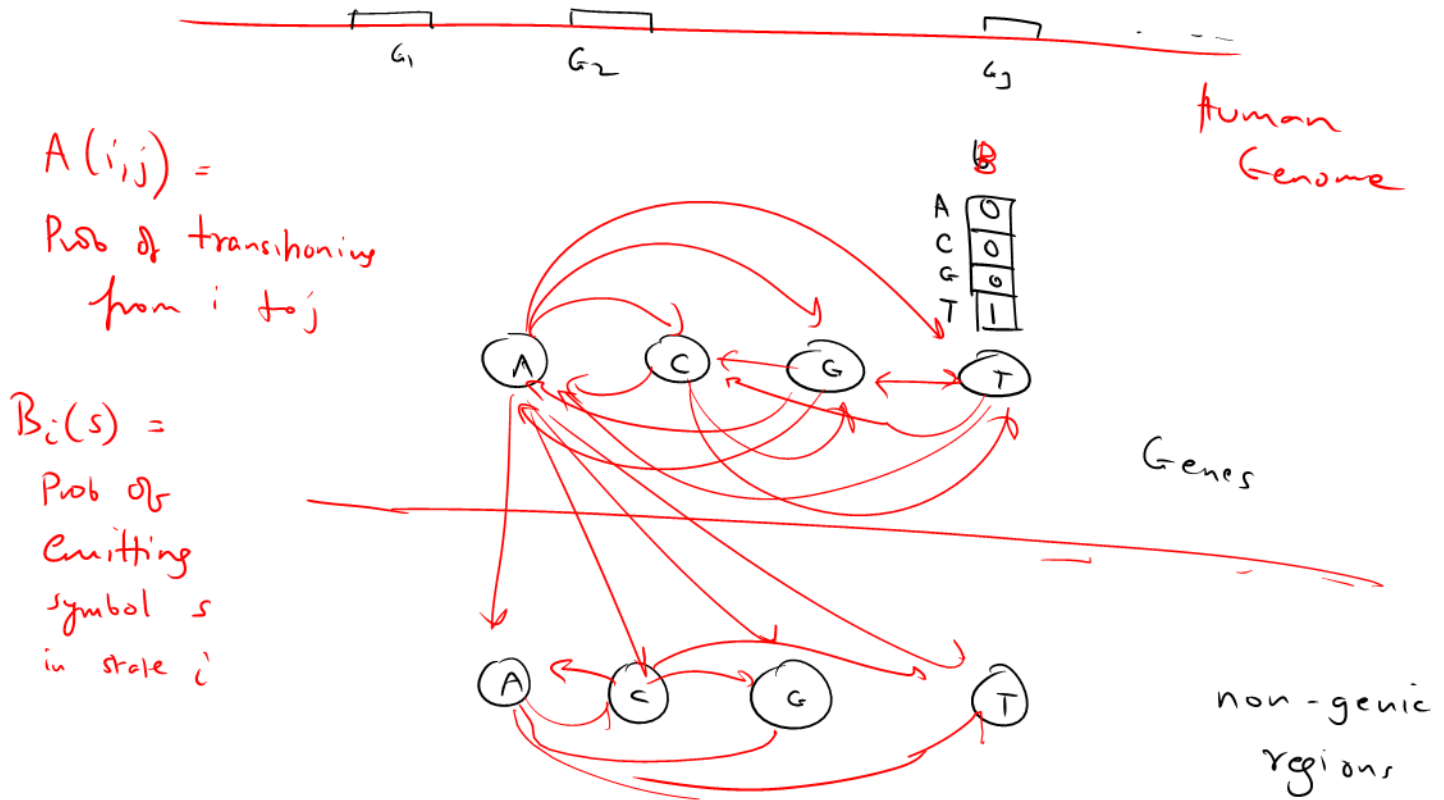


Recap of HMMs

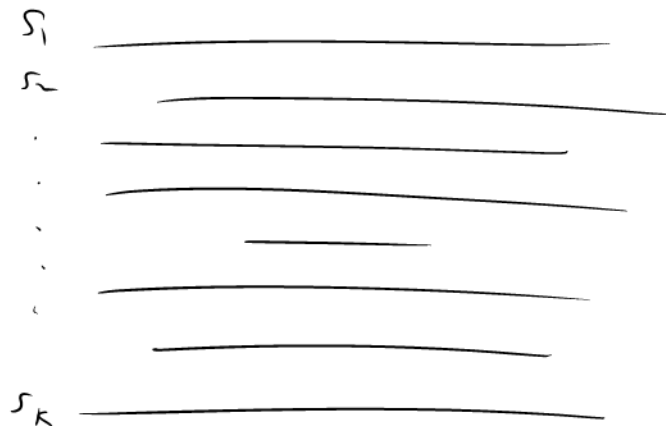


$s = \text{ACAAGCT}$

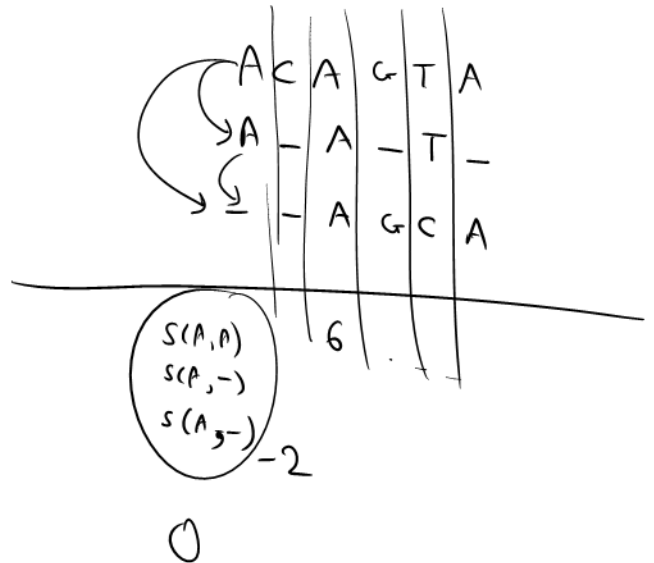
- ① $P(s | \text{model})$
 - ② most like state sequence
 - ③ model training (Baum-welch)
- dynamic programming
- EM Algorithm
Expectation Maximization

Multiple Sequence Alignment & Motif Discovery

average length n

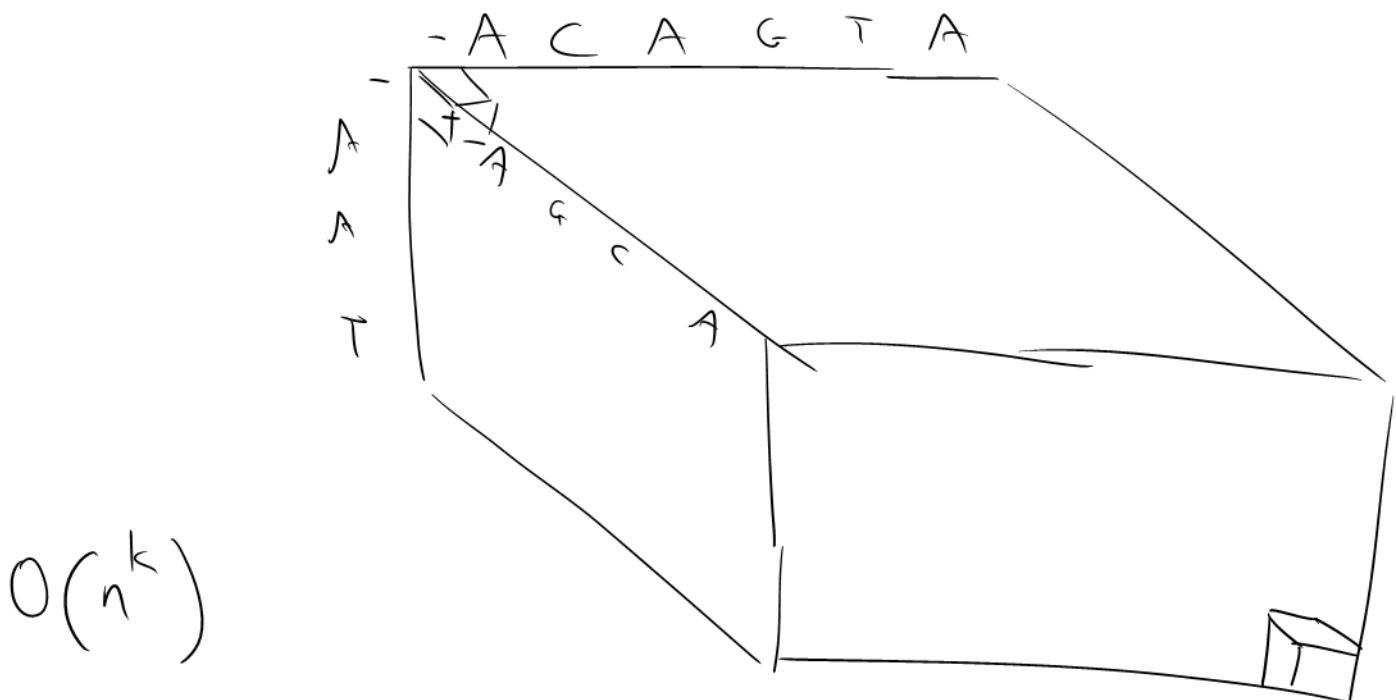


match = 2
mismatch = 1
gap = -1



$$\Sigma : 0 + -2 + 6 + \dots$$

sum of pairs scoring scheme

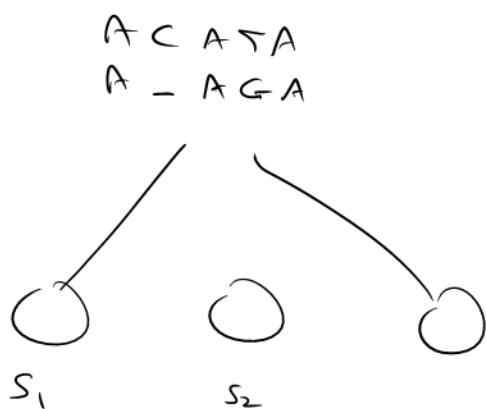
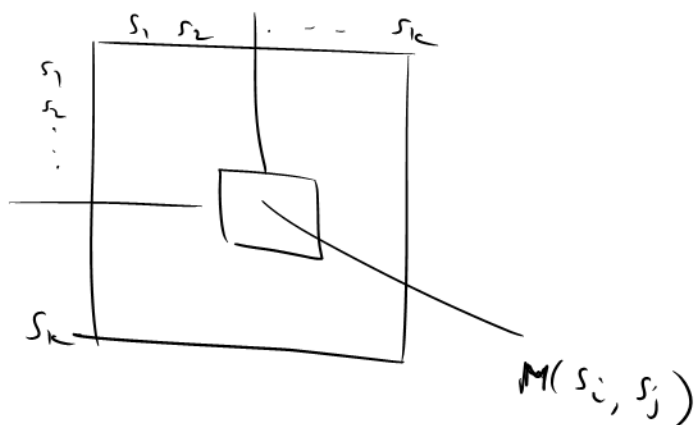


$$O(n^k)$$

s_1 _____
 s_2 _____
 \vdots _____
 \vdots _____
 s_k _____

① Do pairwise alignments

$$\binom{k}{2} O(n^2) \approx O(k^2 n^2)$$



A C A T A
 A - A T A

A - - T A
 - C A G A

X

		-	A	C	A	T	A
		-	A	-	A	T	A
- -	0	-2					
A -							
- C							
- A							
T G							
A A							

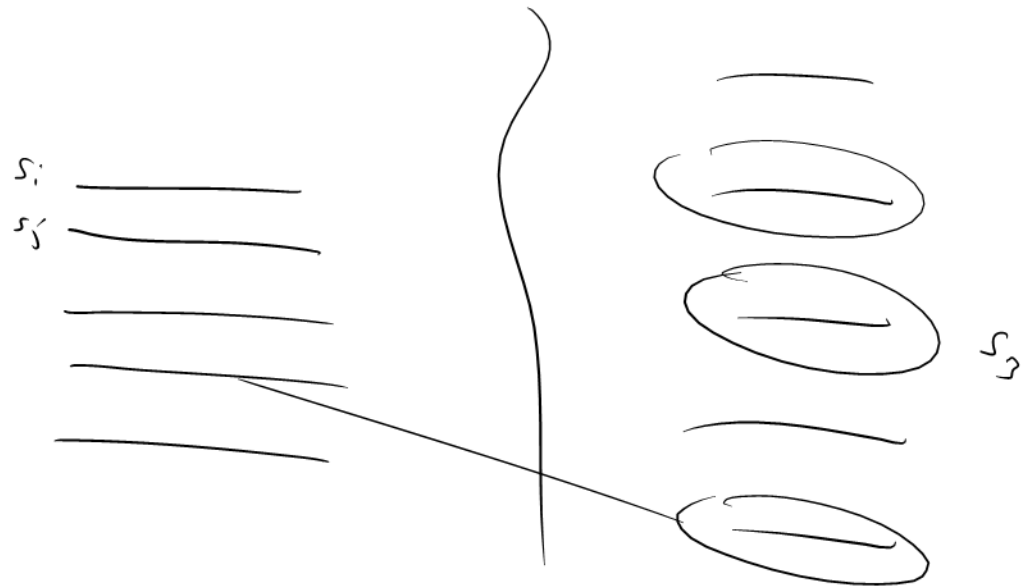
A
 A
 -
 -

Star alignment

①

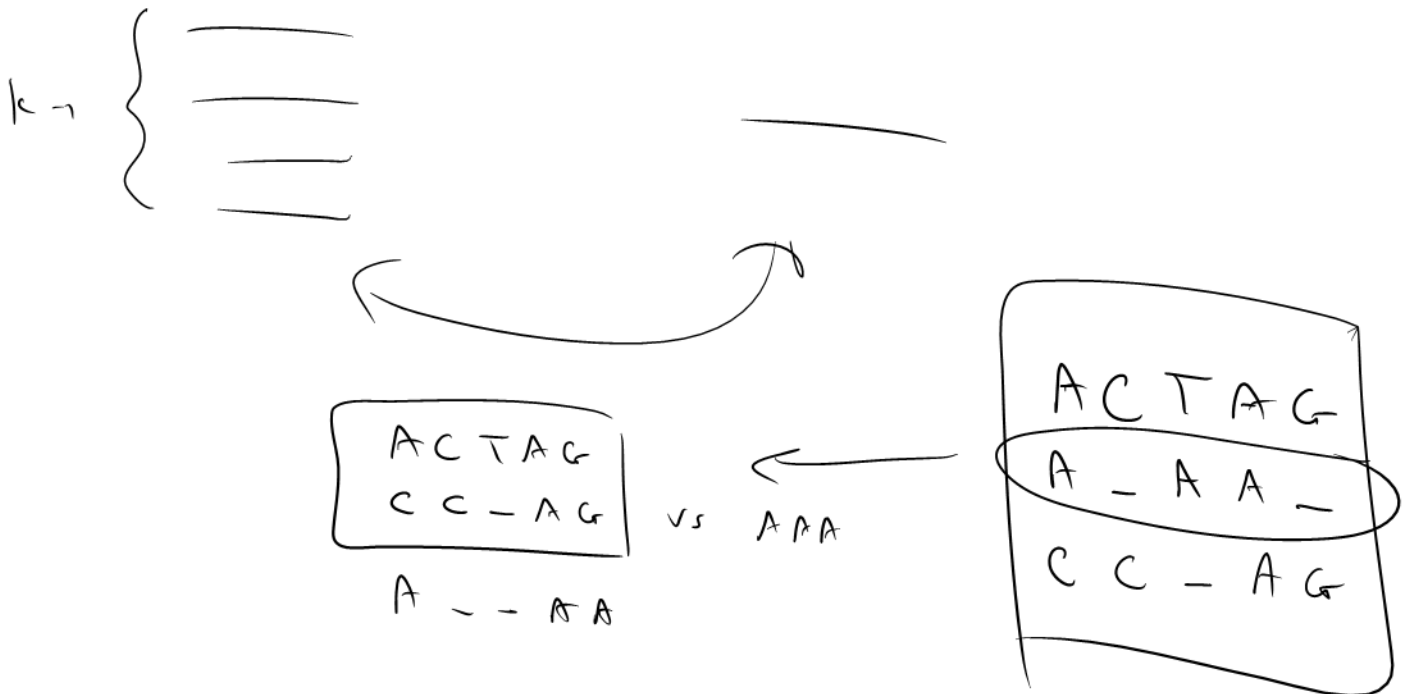
$\binom{k}{2}$

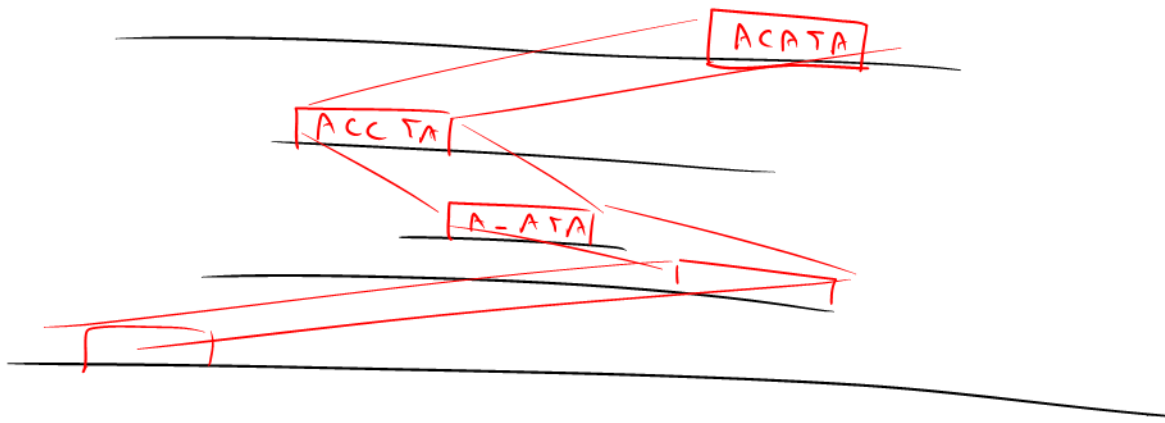
pairwise



③ local refinement

Remove one sequence at random
& realign.





ACATA
ACCTA
A-ATA

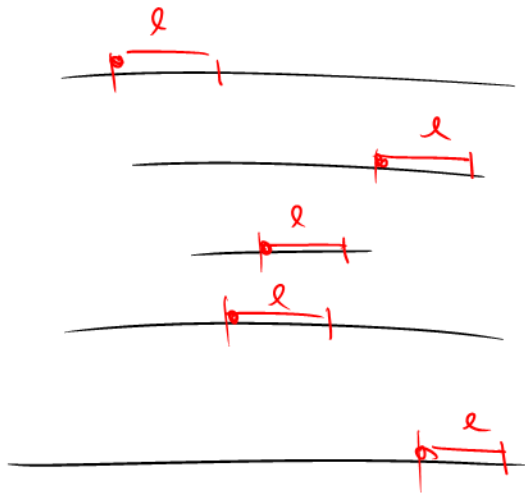
Motif Discovery Problem:

high scoring

find a ^{common} short fragment from at least q sequences
out of k input sequences

Gilbs Sampling

$l \leftarrow$ expected motif length ($q = 100\%$)



PSSM
Position specific Scoring
Matrix

(motif probabilities)

vs

background probabilities

1) Random selection of k positions for the start

↓
A C A T A
C G T
T T C A G A T T

we don't want

0 prob

so adjust by a small amount.

$l = 3$

		1	2	3
Motif	A	0	$\frac{2}{3}$	0
	C	1.0	0	0
	G	0	$\frac{1}{3}$	$\frac{1}{3}$
	T	0	0	$\frac{2}{3}$

background

A	$\frac{3}{7}$
C	0
G	0
T	$\frac{4}{7}$

0.1/1.4
1.1/1.4
0.1/1.4
0.1/1.4

e.g. add 0.1 to all entries

prior probability

