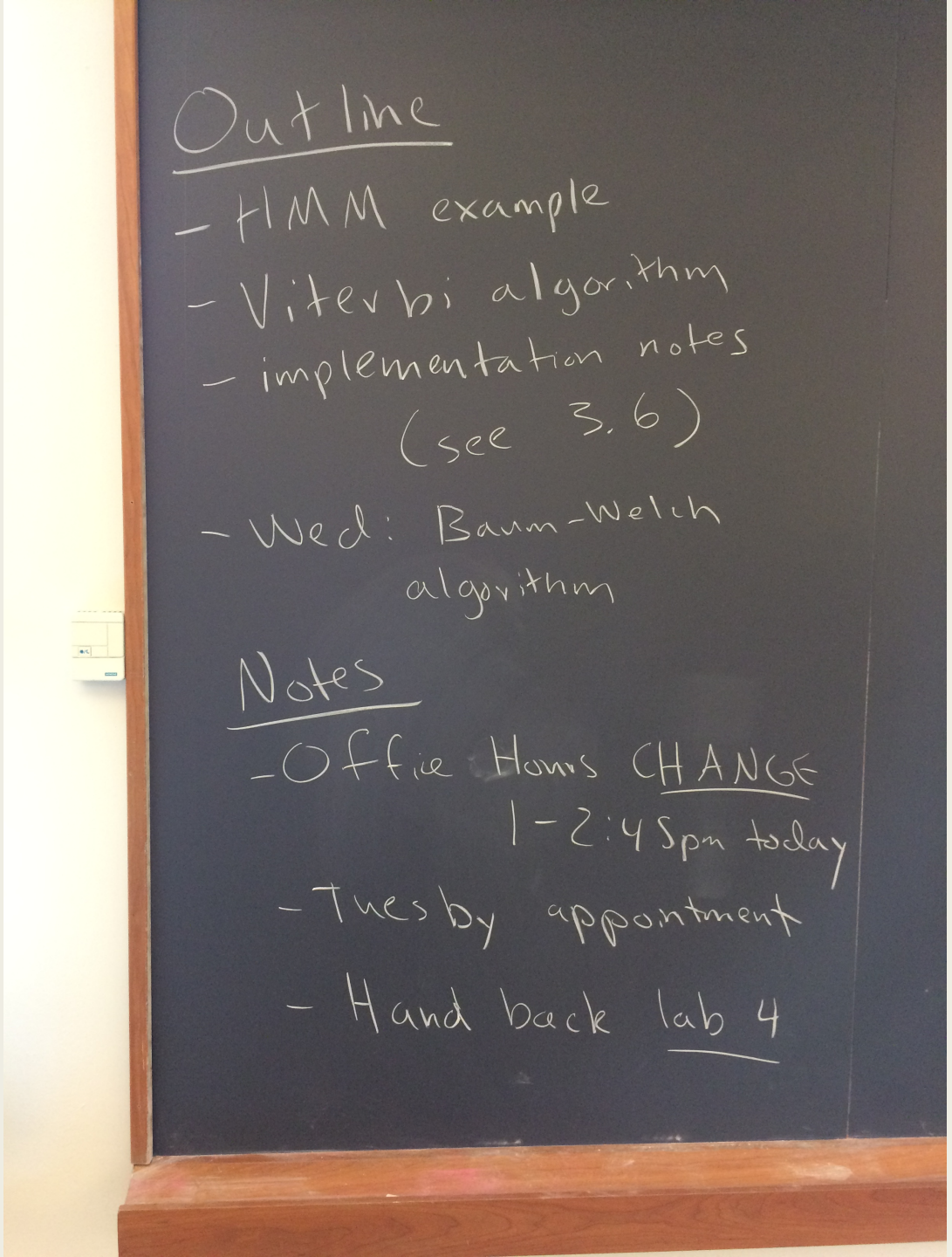
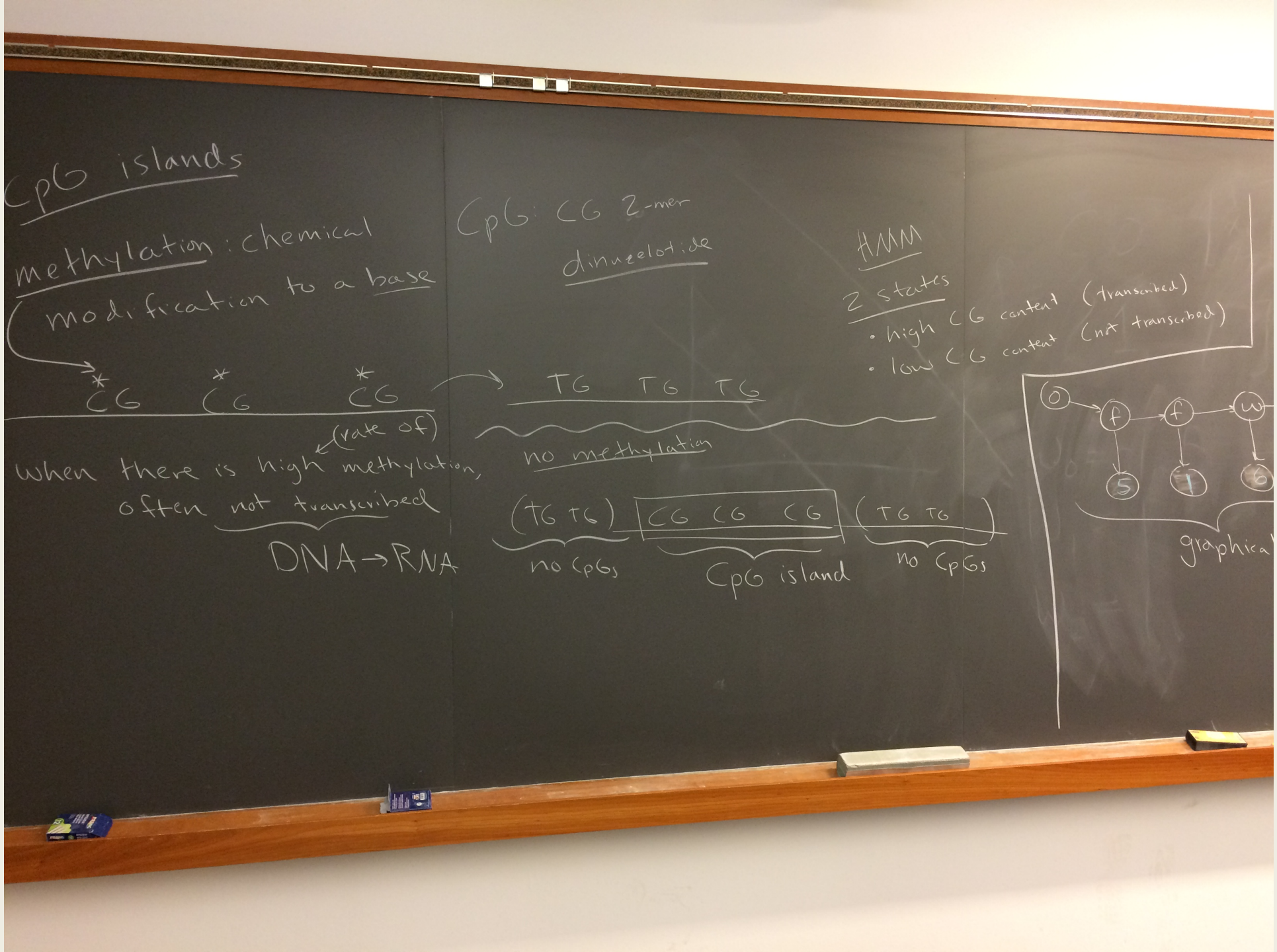
CS 68: BIOINFORMATICS

Prof. Sara Mathieson
Swarthmore College
Spring 2018

Apr 2


Outline
- HMM example
- Viterbi algorithm
- implementation notes (see 3.6)
- Wed: Baum-Welch algorithm
Notes
- Office Hours CHANGE 1-2:45pm today
- Thurs by appointment
- Hand back lab 4

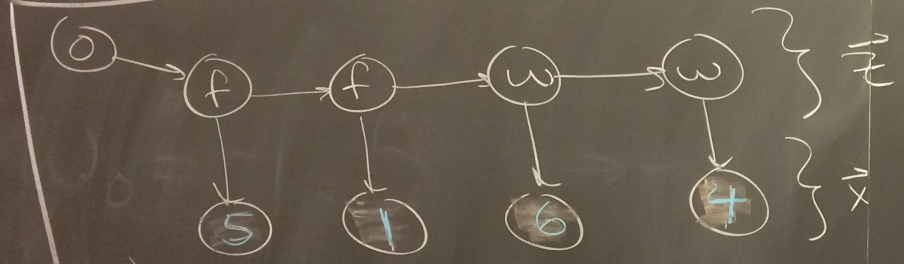
HMM Example: CpG Islands


CpG islands
methylation: chemical modification to a base
CpG: CG 2-mer dinucleotide
HMM
2 states
• high CG content (transcribed)
• low CG content (not transcribed)
CG CG CG
TG TG TG
no methylation
(rate of)
when there is high methylation, often not transcribed
DNA→RNA
(TG TG)
CG CG CG
(TG TG)
no CpGs
CpG island
no CpGs
0
f
f
w
5
1
6
graphica

Viterbi Algorithm

ontent (transcribed)

ontent (not transcribed)



graphical model

HMM definition

$a_{kl} = P(z_i = l \mid z_{i-1} = k)$

(transition probabilities)

$e_k(b) = P(x_i = b \mid z_i = k)$

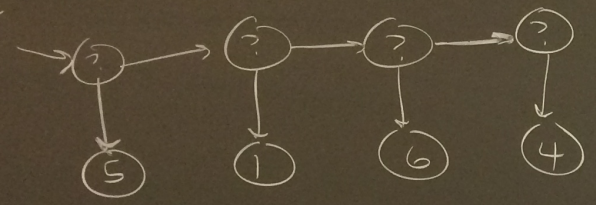
joint probability of $\vec{x}$ & $\vec{z}$

$$P(\vec{x}, \vec{z}) = \prod_{i=1}^{L} a_{z_{i-1} z_i} e_{z_i}(x_i)$$

(if we knew $\vec{z}$)

Q: what is the most likely hidden state sequence $\vec{z}^*$?

A: Viterbi algorithm (DP)



K = # of hidden states (K=2)

B = # of possible emissions (B=6)

L = length of sequence (both $\vec{x}$ & $\vec{z}$)

naive

possible hidden state sequences:

(1) $O(K^L)$ ← try all & take max $P(\vec{x}, \vec{z})$

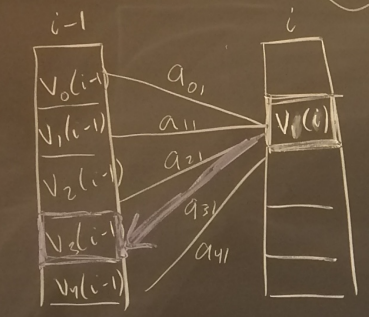
Viterbi recursion

$V_k(i)$ = prob of the best path that ends at x_i with state k

(2)

Handout 23

$$V_k(i) = e_k(x_i) \cdot \max_l \{ V_l(i-1) \cdot a_{lk} \}$$

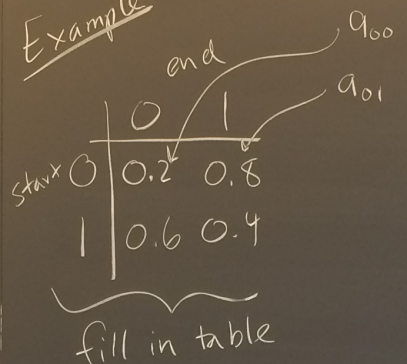


states $= \{0, 1, 2, 3, 4\}$

k

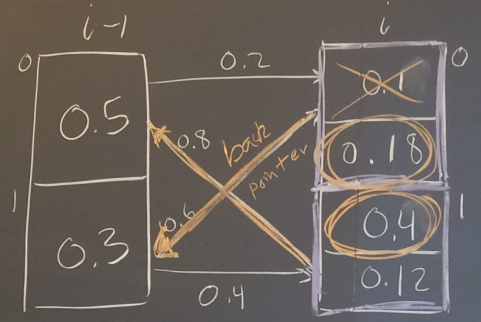
② What is the runtime of Viterbi? (K, B, L)?

Example



fill in table

Markov chain transition probabilities

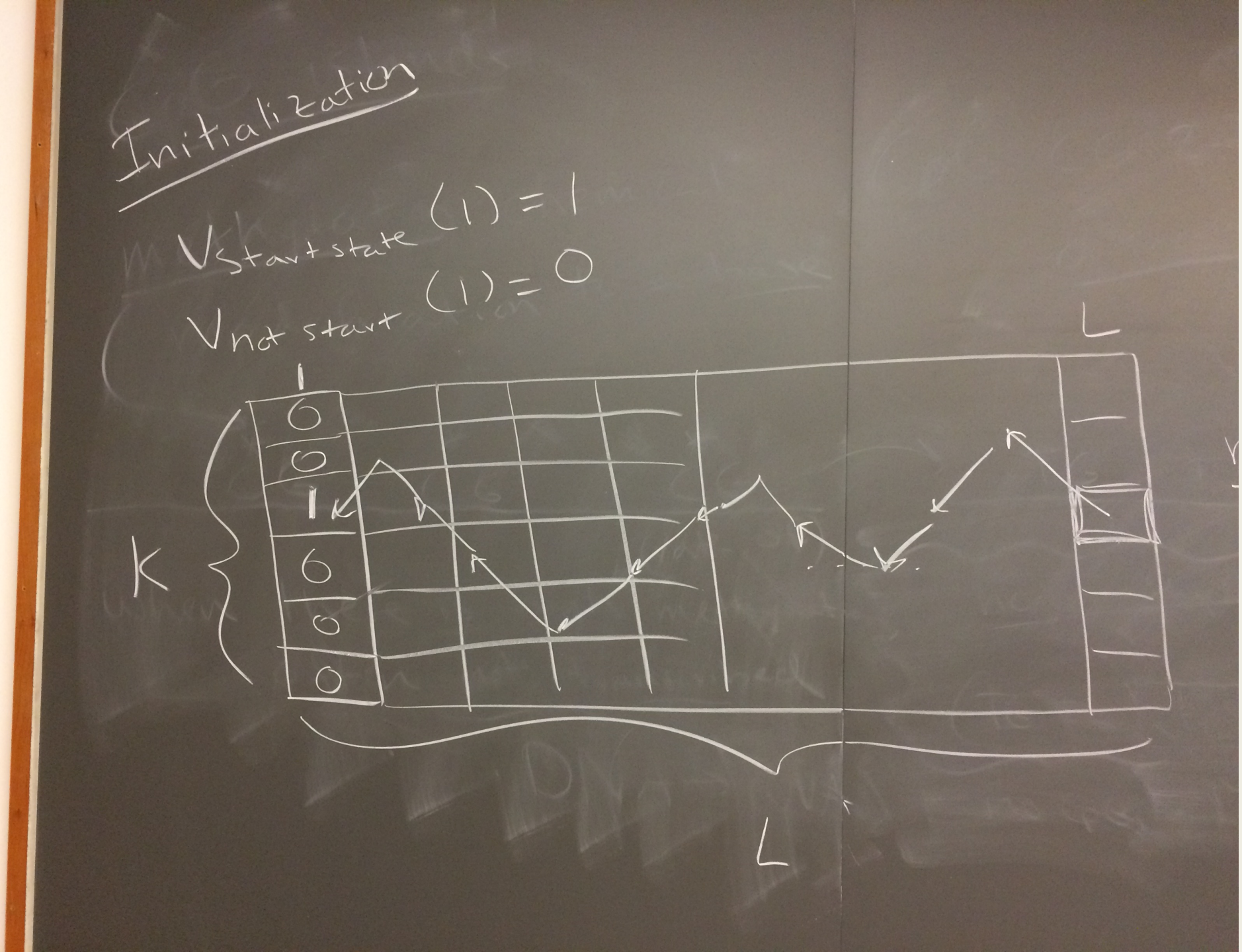


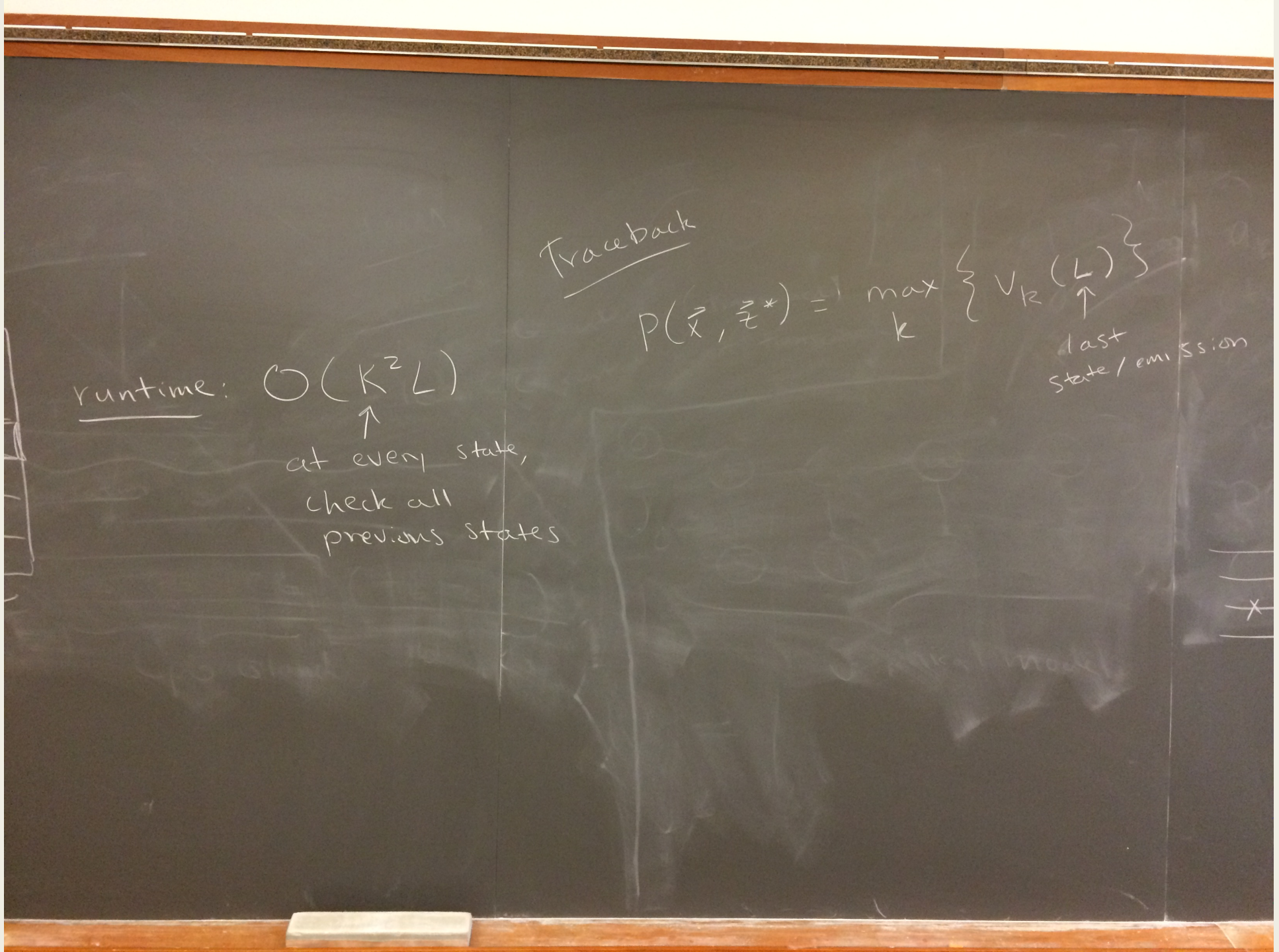
$V_0(i-1) = 0.5$

$V_1(i-1) = 0.3$

fill in the arrows

assume all emissions equally likely


Initialization
$V_{start\ state}(1) = 1$
$V_{not\ start}(1) = 0$
L
1
0
0
1
0
0
0
K
L


Traceback
$P(\vec{x}, \vec{z}^*) = \max_k \{ V_k(L) \}$
last state / emission
runtime: $O(K^2 L)$
at every state, check all previous states