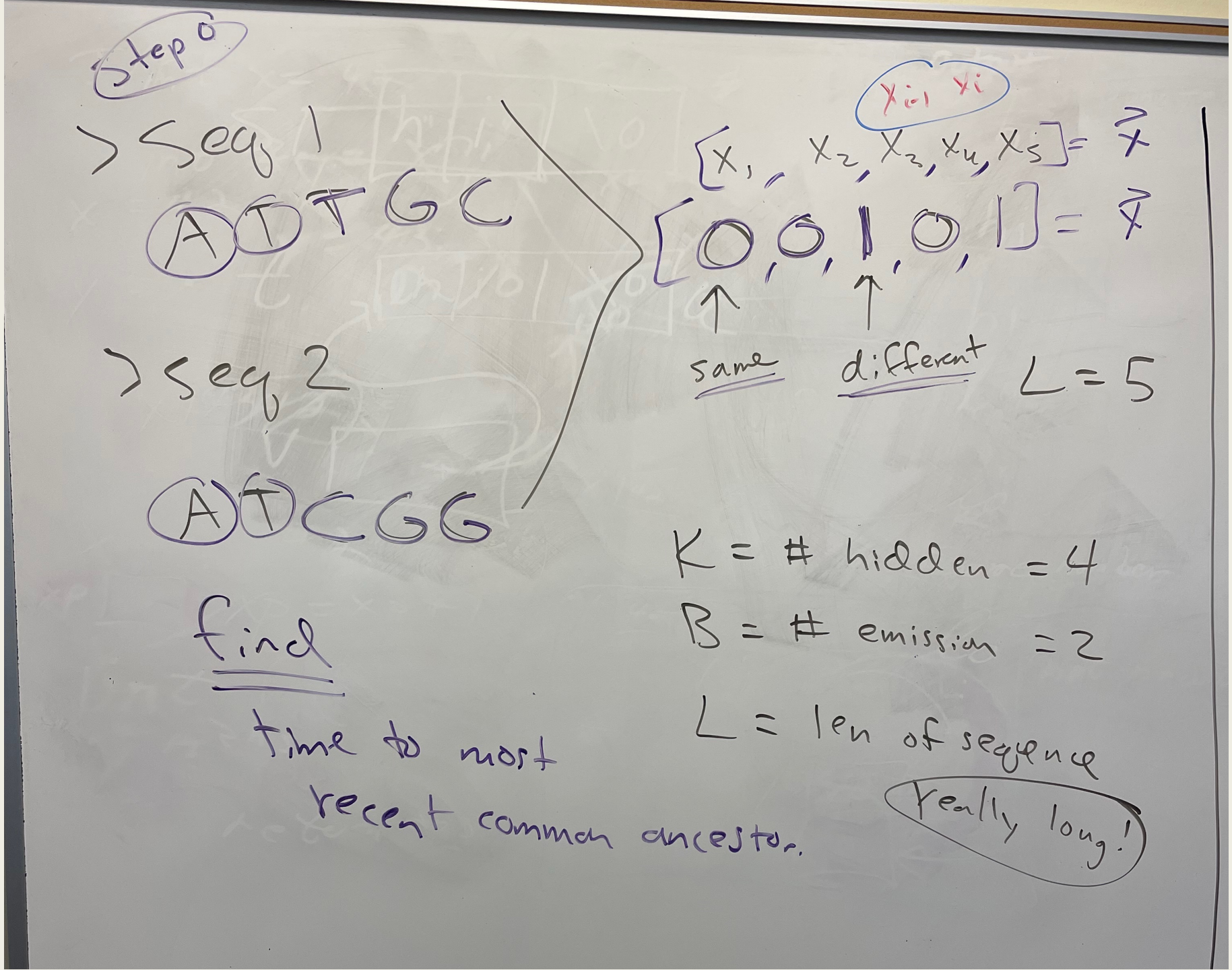
CS 364 COMPUTATIONAL BIOLOGY

Sara Mathieson
Haverford College

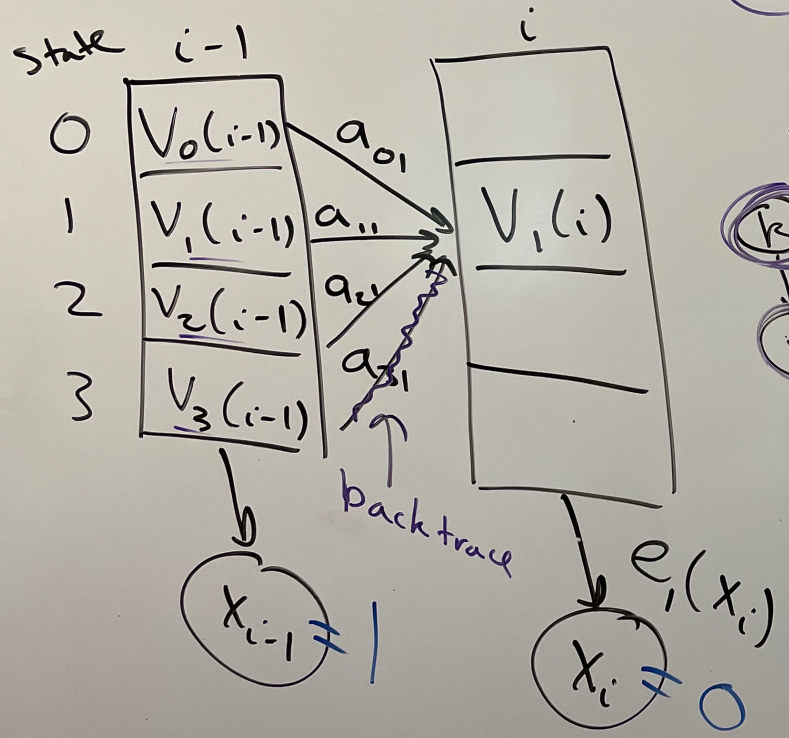
In-lab notes for Lab 8

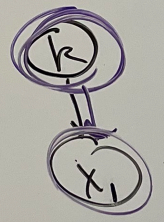

step 0
>seq 1
ATTGC
>seq 2
ATCGG
find
time to most recent common ancestor.
x_{i-1} x_i
$[x_1, x_2, x_3, x_4, x_5] = \vec{x}$
$[0, 0, 1, 0, 1] = \vec{x}$
same
different
$L = 5$
K = # hidden = 4
B = # emission = 2
L = len of sequence
really long!

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

$= e_k(x_i) \cdot \max_{\ell} \{ V_\ell(i-1) \cdot a_{\ell k} \}$

$a_{\ell k}$ ↑ transition from state $\ell \to k$





initialization

$V_k(1) = \pi_k e_k(x_1)$

prob of starting in state k ; first obs

π_k $k=0$ $x_1=0$

$\log \pi_k + \log e_k(x_1)$

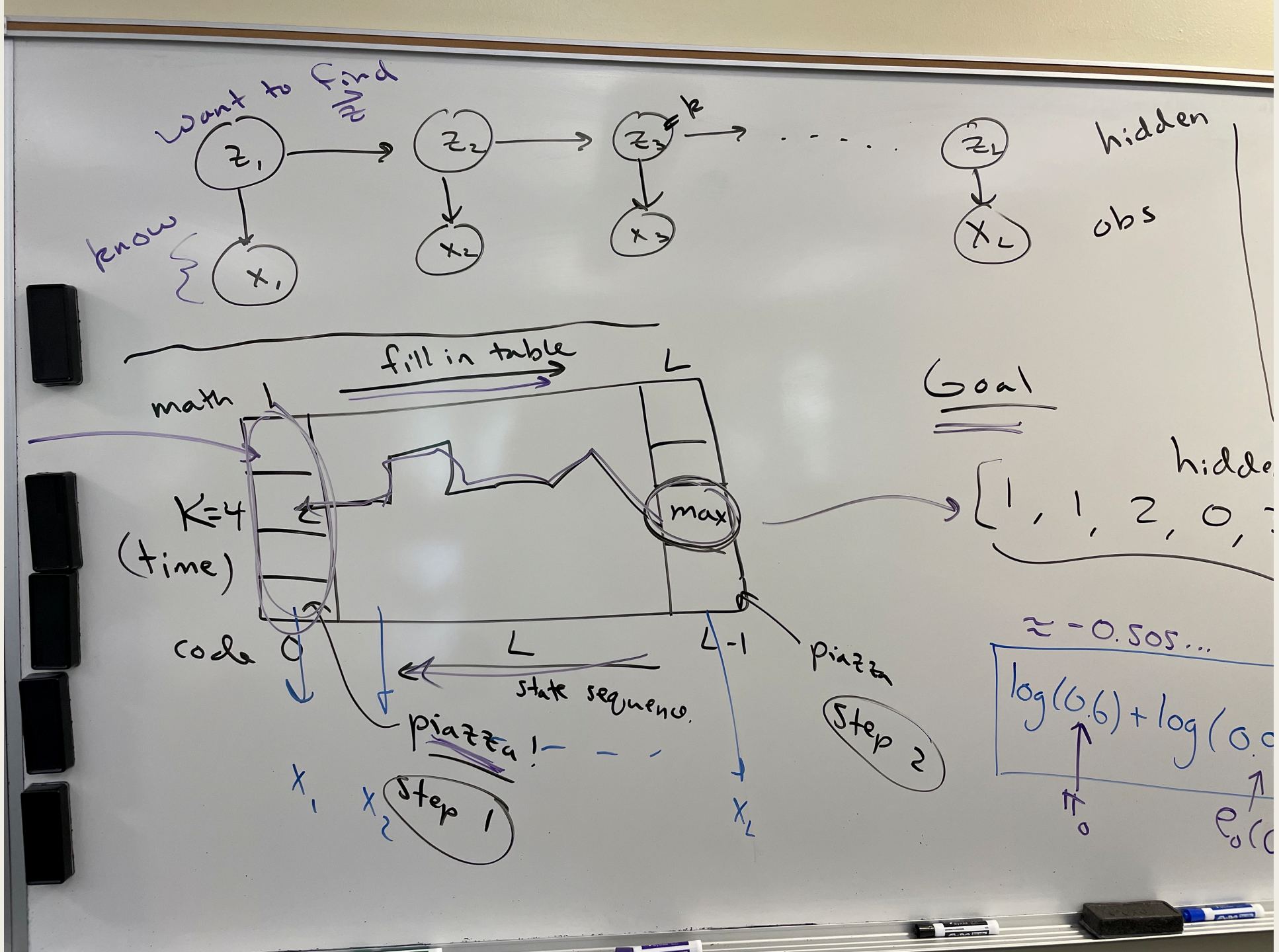
termination

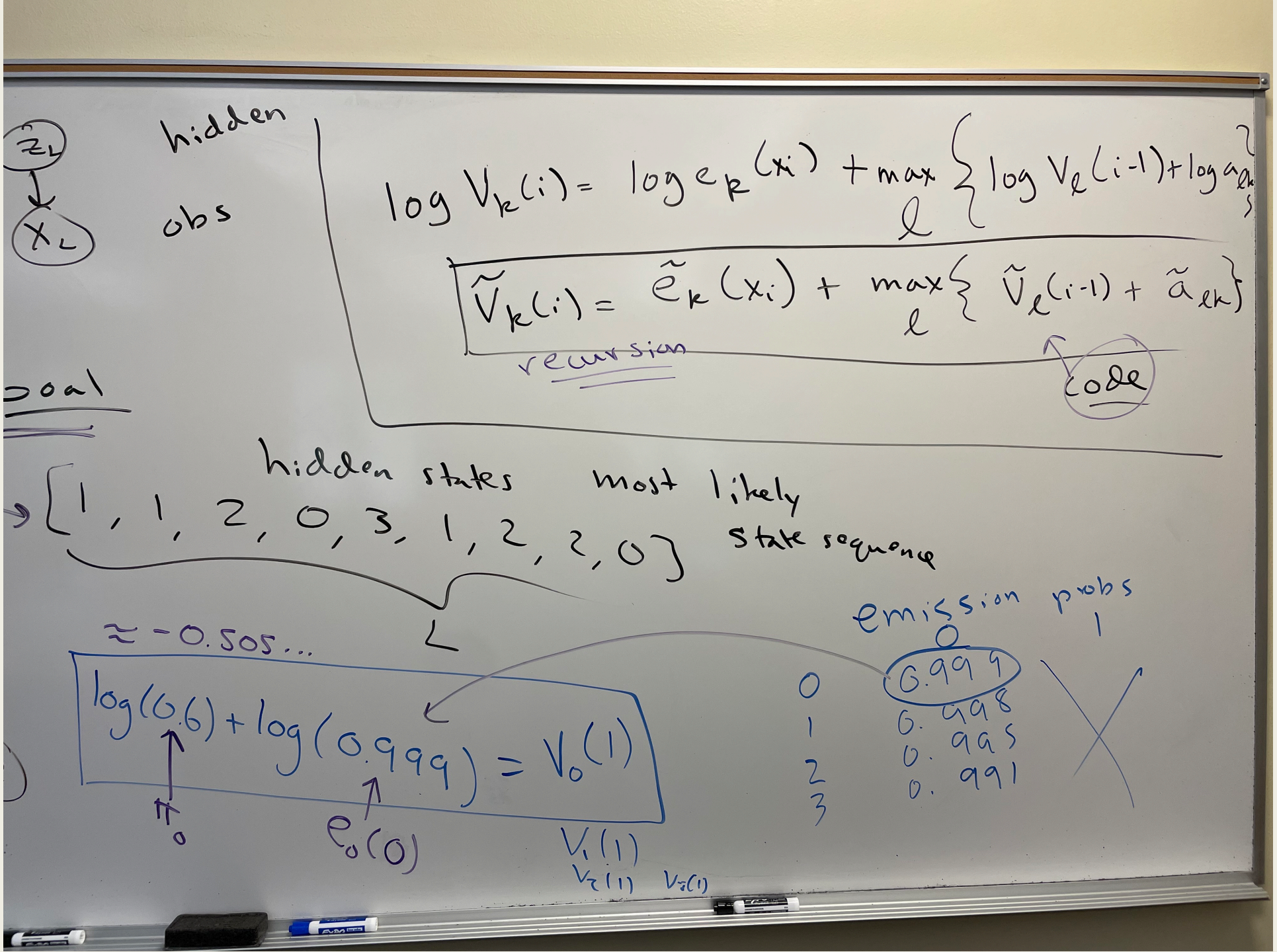
$p(\vec{x}, \vec{z}^*) =$

$\vec{z}^*$ ↑ best path

$\max_k \{ V_k(L) \}$

last column of DP table


Want to find
z_1
z_2
z_3
k
z_L
hidden
know
x_1
x_2
x_3
x_L
obs
fill in table
math
L
K=4
(time)
max
code
0
L
L-1
piazza
state sequence
piazza!
Step 1
Step 2
Goal
hidde
[1, 1, 2, 0,
≈ -0.505...
log(0.6) + log(0.
π_0
e_0(


hidden
obs
$\log V_k(i) = \log e_k(x_i) + \max_l \{\log V_l(i-1) + \log a_{lk}\}$
$\tilde{V}_k(i) = \tilde{e}_k(x_i) + \max_l \{\tilde{V}_l(i-1) + \tilde{a}_{lk}\}$
recursion
code
goal
hidden states
most likely state sequence
[1, 1, 2, 0, 3, 1, 2, 2, 0]
≈ -0.505...
$\log(0.6) + \log(0.999) = V_0(1)$
π_0
$e_0(0)$
$V_1(1)$
$V_2(1)$ $V_3(1)$
emission probs
0
1
2
3
0.999
0.998
0.995
0.991