Genome 540 Discussion

February 27th, 2024

Clifford Rostomily

Genome Sciences
UNIVERSITY OF WASHINGTON

Assignment 7 Questions?

- Part 1: Use your predicted D-segments from hw6 to
  - Generate a new scoring scheme
  - Simulate background sequence
- Part 2: Run your D-segment program on the background and compare to the real data
- Part 3: Answer some questions

Assignment 8

HMM Tasks

Rabiner 1989:

Likelihood: Given an HMM λ = (A, B) and an observation sequence O, determine the *likelihood* P(O|λ).

Decoding: Given an observation sequence O and an HMM λ = (A, B), discover the *best hidden state* sequence Q.

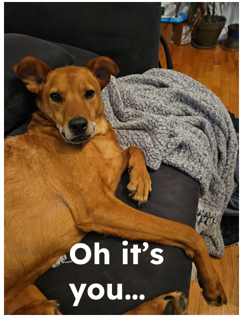
Learning: Given an observation sequence O and the set of states in the HMM, learn the HMM *parameters A and B.*

Example

Your dog is very moody and you want to know when they **like** or **hate** you so you start recording what they are doing when you get home everyday…

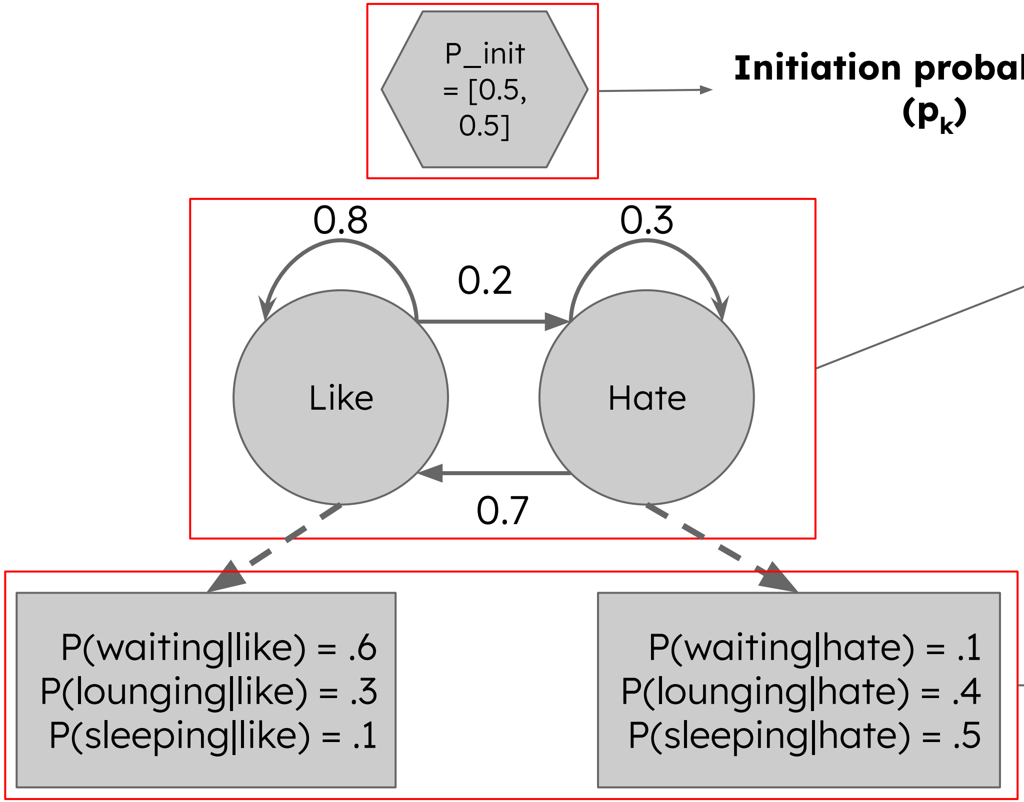
Waiting

Lounging



Sleeping

Model



Initiation probabilities (p_k)

Transition Matrix (A)

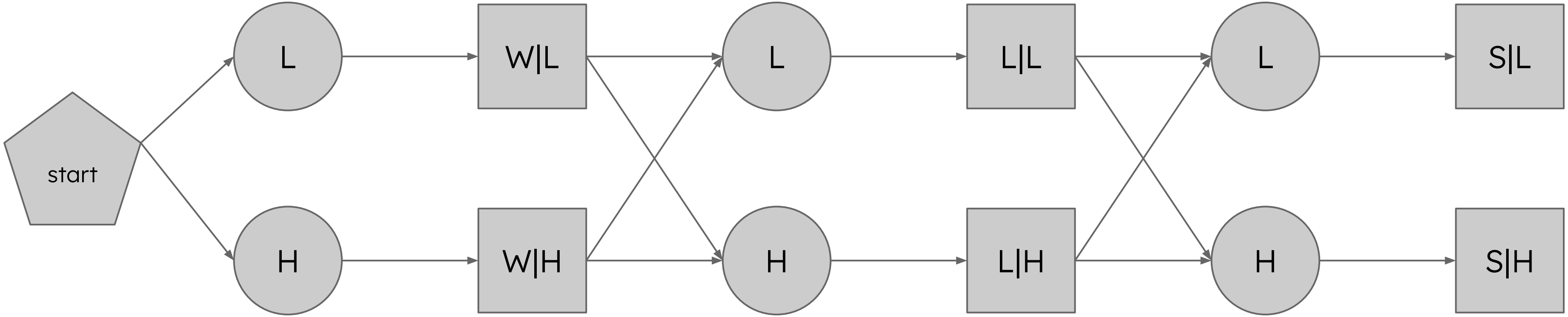
	L	H
L	0.8	0.2
H	0.7	0.3

a_{LH}

Set of Hidden States (S = {k})

Emission Probabilities of the observations (B = {b}) given the state

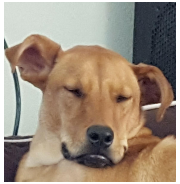
Graphical representation with data



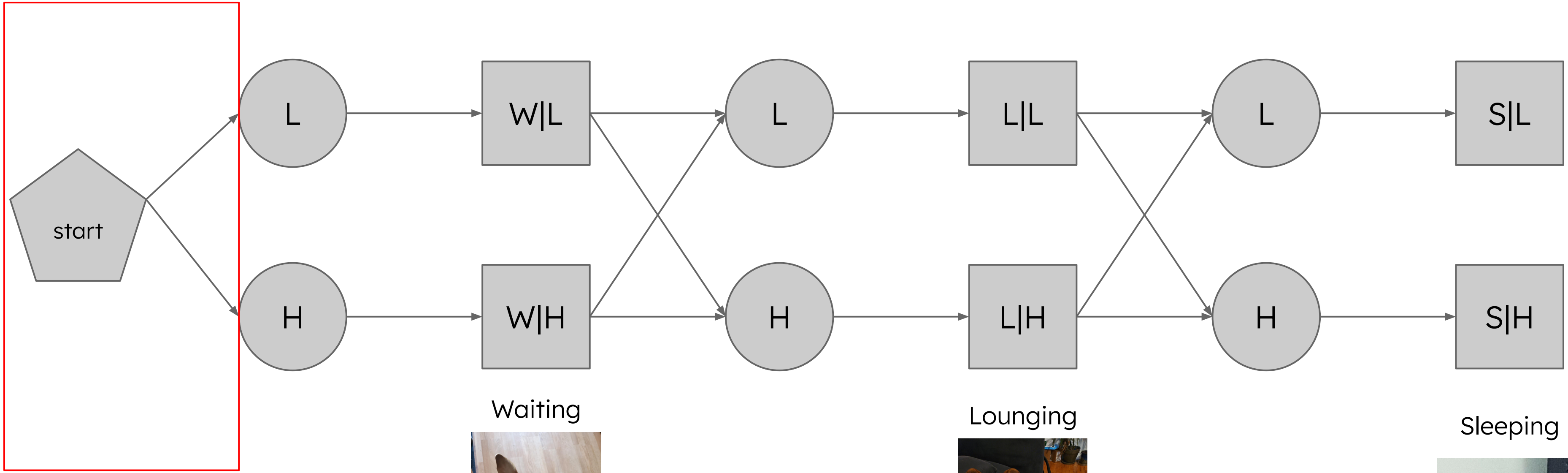
Waiting

Lounging

Sleeping



Graphical representation with data

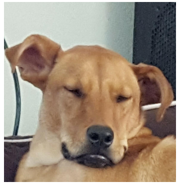


Initiation

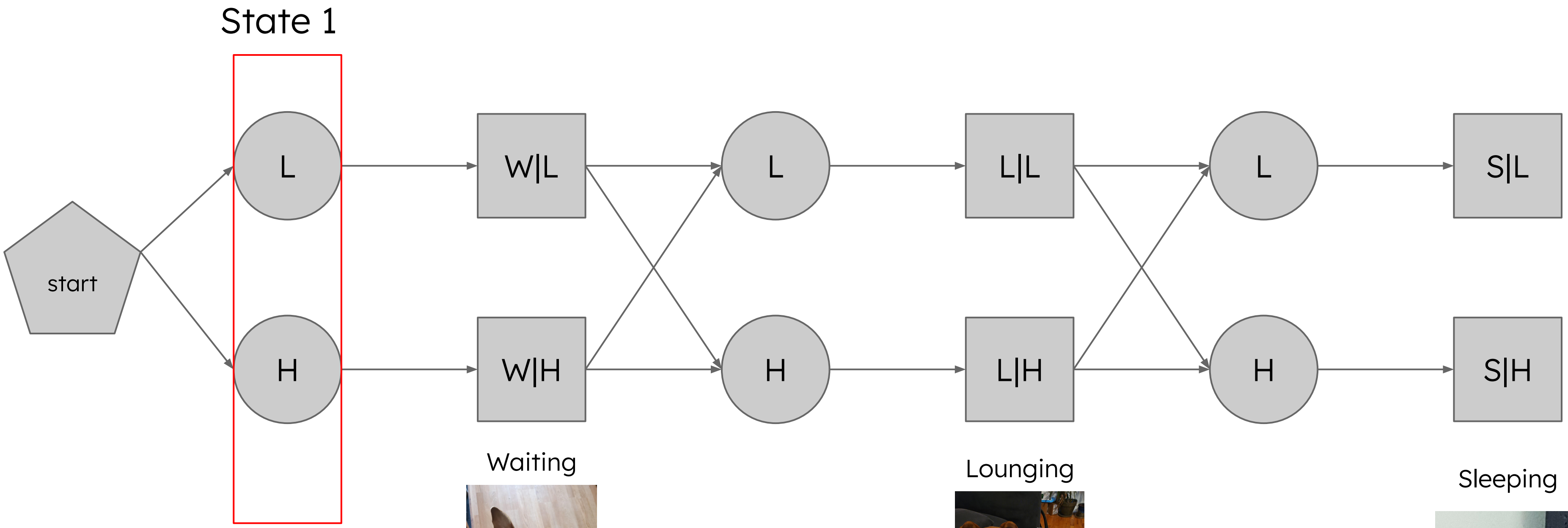
Waiting

Lounging

Sleeping



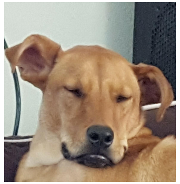
Graphical representation with data



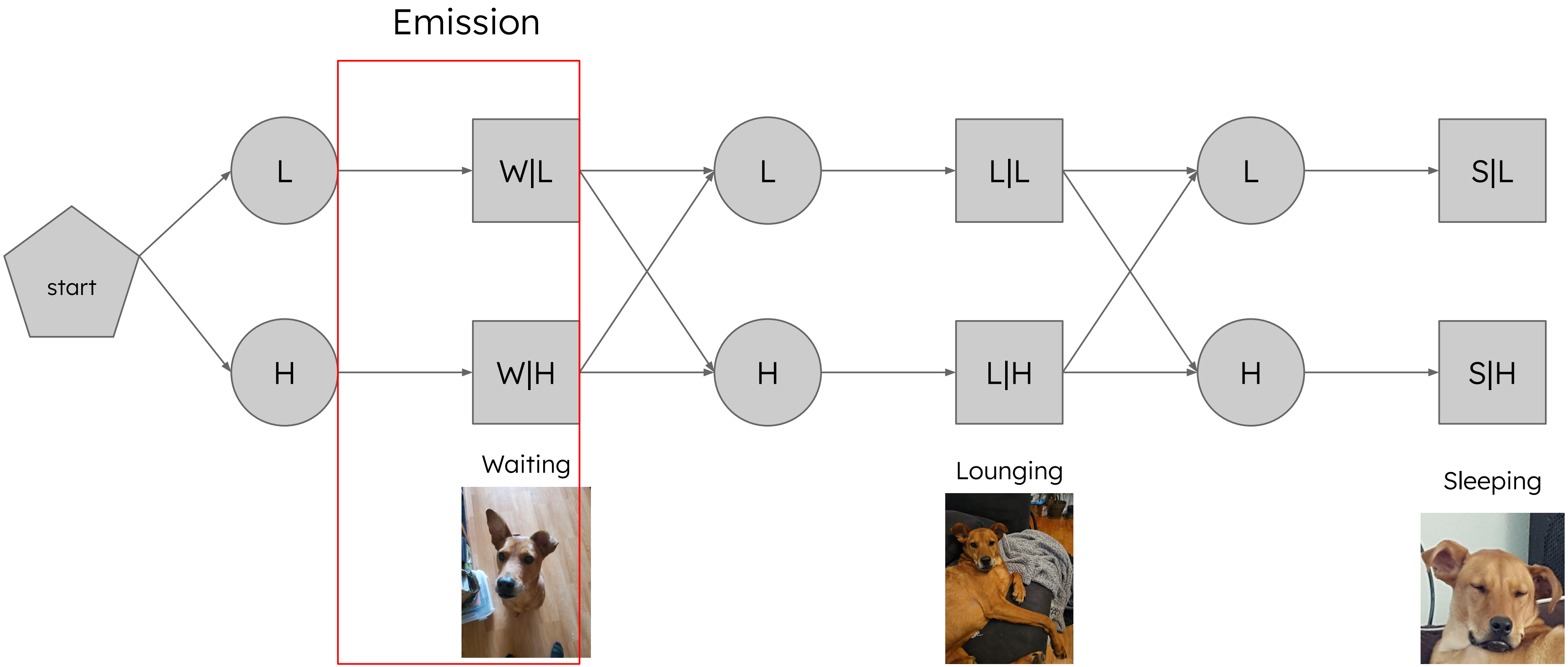
Waiting

Lounging

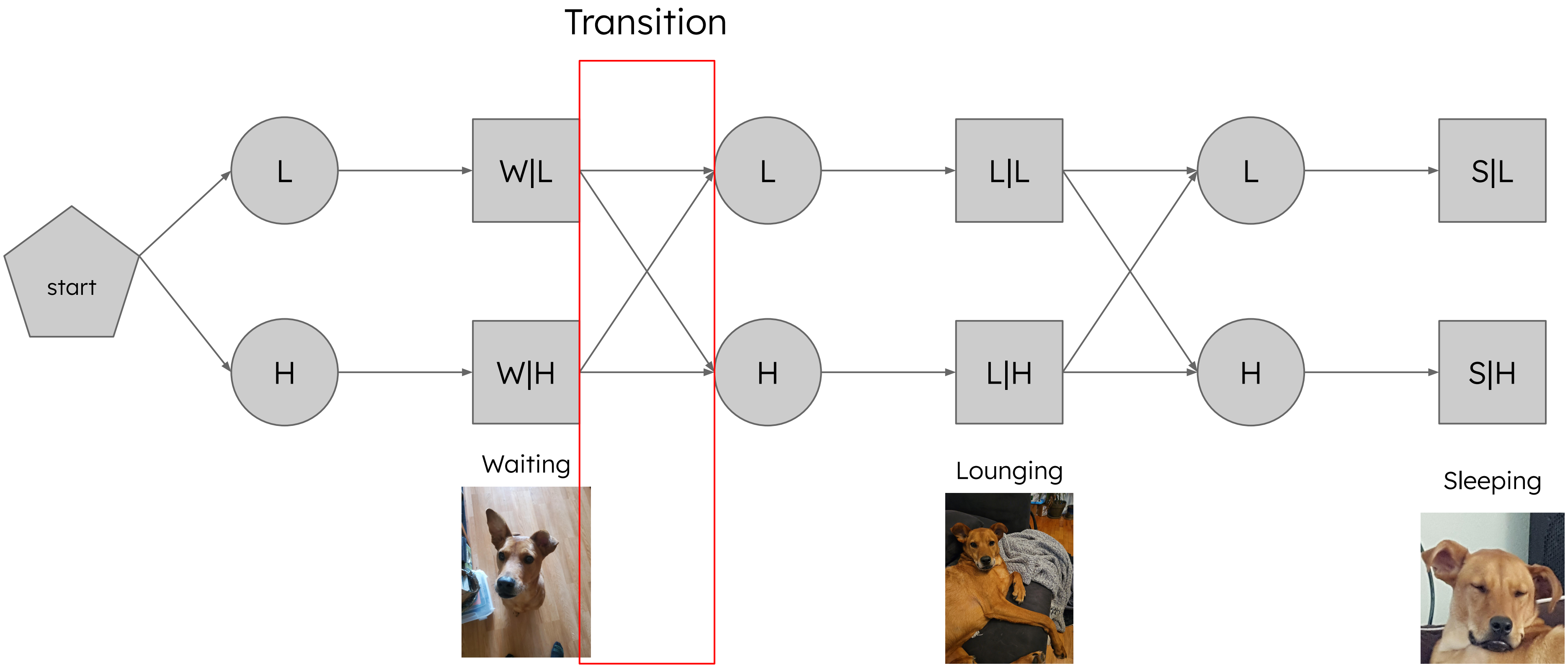
Sleeping



Graphical representation with data



Graphical representation with data



Baum Welch (Forward/Backward) - “Training” an HMM

1. Step 1: Expectation
   a. Compute the forward probabilities
   b. Compute the backward probabilities
2. Step 3: Maximization
   a. Update the transition and emission probabilities

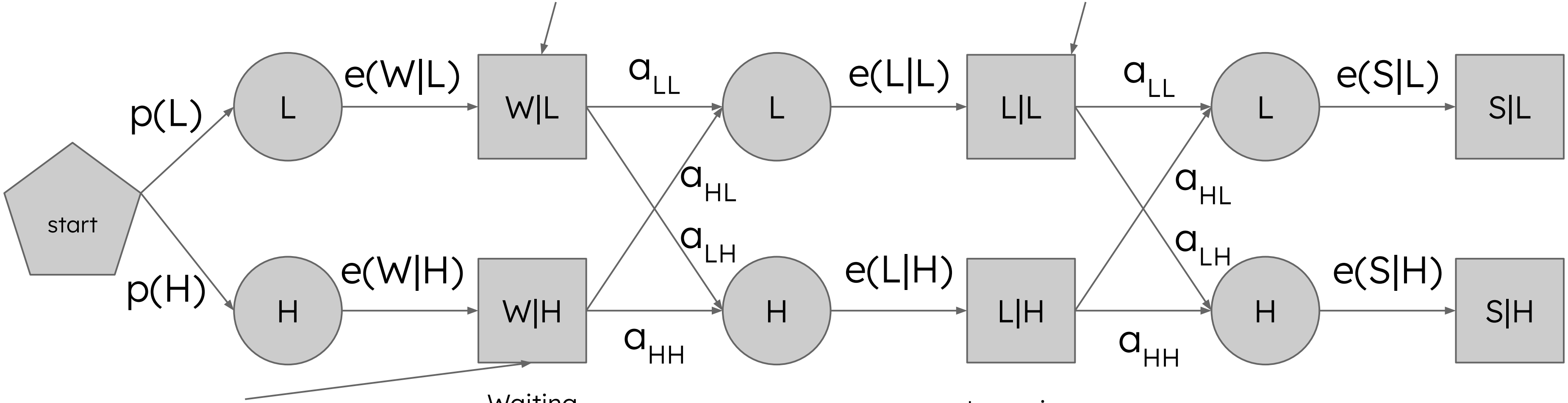
Forward Algorithm - **Likelihood** of an observed sequence

3 steps:

1. Initialization
2. Recursion
3. Termination

Forward Algorithm - **Likelihood** of an observed sequence

$f_1(L)=p(L)*e(W|L)$ $f_2(L)=(f_1(L)*a_{LL} + f_1(H)*a_{HL})*e(L|L)$



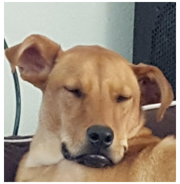
$f_1(H)=p(H)*e(W|H)$

***Emission may also be written $e_H(W)$

Waiting

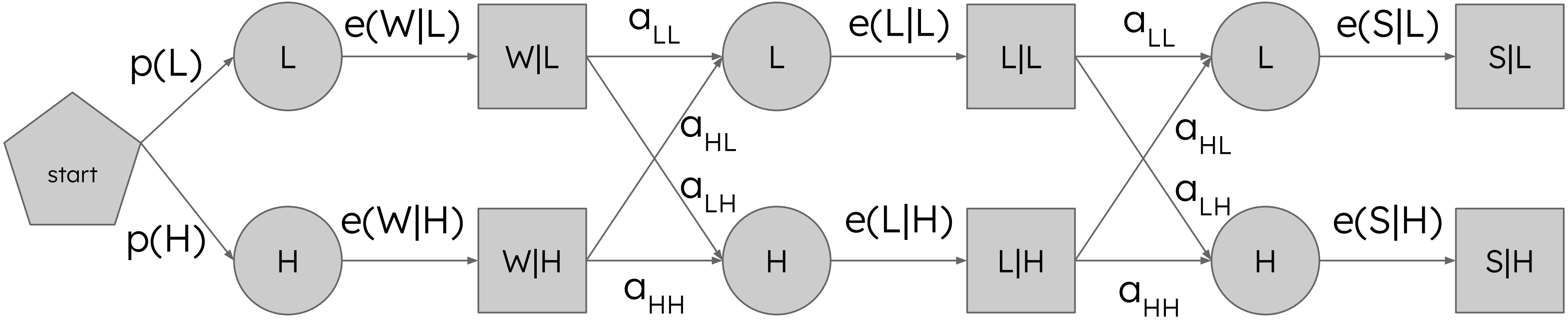
Lounging

Sleeping



Computing the backward probabilities

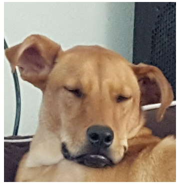
Backward probabilities: probability of seeing the observations from time t + 1 to the end



Waiting

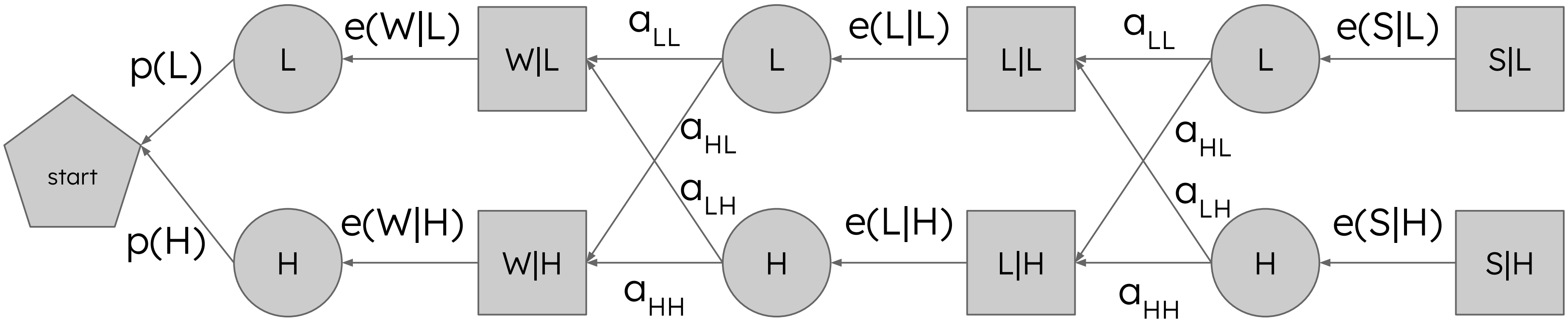
Lounging

Sleeping



Computing the backward probabilities

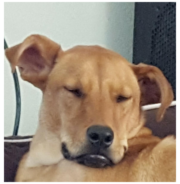
Backward probabilities: probability of seeing the observations from time t + 1 to the end



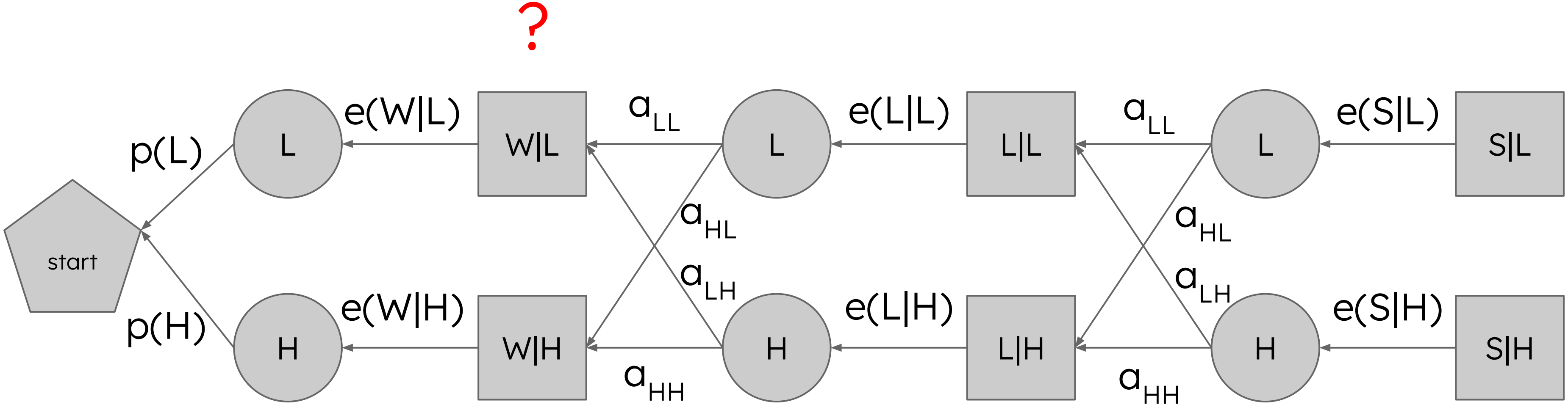
Waiting

Lounging

Sleeping



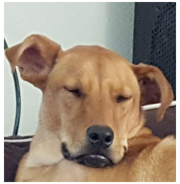
Computing the backward probabilities



Waiting

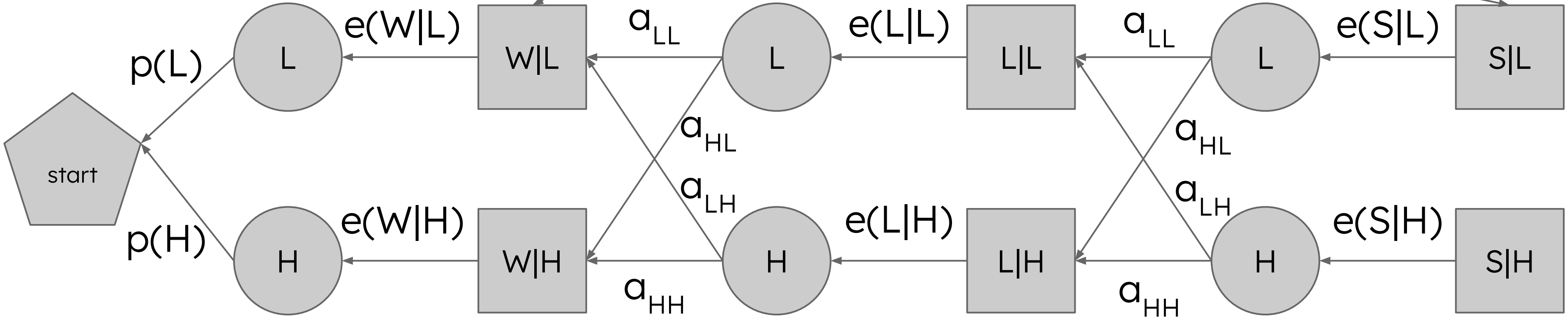
Lounging

Sleeping



Computing the backward probabilities

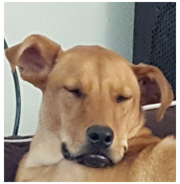
$b_t(i) = b_{t+1}(L)*a_{LL}*e(L|L) + b_{t+1}(H)*a_{LH}*e(L|H)$ **Initialize assuming $b_T(i) = 1$



Waiting

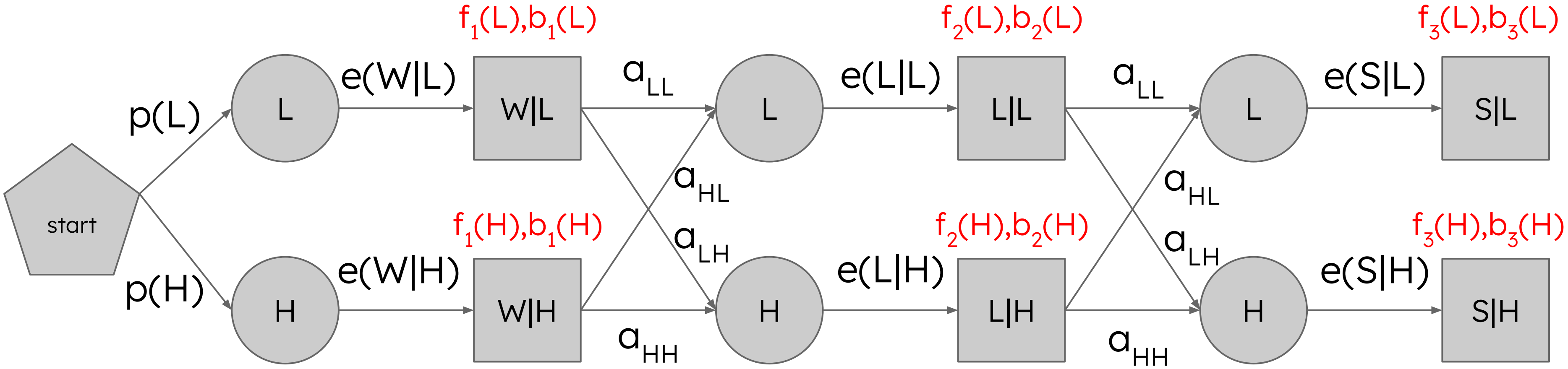
Lounging

Sleeping



$b_t(i)=\Sigma_j b_{t+1}(j)*a_{ij}*e_j(t+1)$

Calculating the transition probabilities

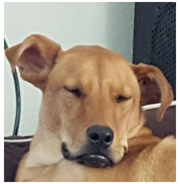


$$P_t(i,j) = \frac{f_t(i) * a_{ij} * e_j(o_{t+1}) * b_{t+1}(j)}{\Sigma^{N}_{j=1} f_t(j)b_t(j)}$$

Waiting

Lounging

Sleeping



Calculating the transition probabilities

$$P_t(i,j) = \frac{f_t(i) * a_{ij} * e_j(o_{t+1}) * b_{t+1}(j)}{\Sigma^{N}_{j=1} f_t(j) b_t(j)}$$

Probability of observations constrained on a specific transition (numerator)

Probability of observations given the model (denominator)

$$\underline{a}(i,j) = \frac{\Sigma^{T-1}_{t=1} P_t(i,j)}{\Sigma^{T-1}_{t=1} \Sigma^{N}_{k=1} P_t(i,k)}$$

Calculating the emission probabilities

$$\gamma_t(j) = \frac{f_t(j)b_t(j)}{P(O|\lambda)} = \frac{f_t(j)b_t(j)}{\Sigma^{N}_{j=1} f_t(j)b_t(j)}$$

← Probability of being in state j at time t given the observation sequence O and the model

$$e_t(v_k|j) = \frac{\text{expected number of times in state j and observing symbol } v_k}{\text{expected number of times in state j}}$$

$$e_t(v_k|j) = e_j(v_k) = \frac{\Sigma^{T}_{t=1,Ot=vk}\gamma_t(j)}{\Sigma^{T}_{t=1}\gamma_t(j)}$$

← Sum of all $\gamma_t(j)$ where the observed symbol = v_k

Avoiding vanishing probabilities

- Scaling
  - Good tutorial
- Work in log space
  - Mann 2006

Scaling

- When computing forward probabilities, also compute a scaling factor c_t

$$c_t = \frac{1}{\sum^{N}_{i=1} f_t(i)}$$

- New forward probabilities at time t are multiplied by c_t
- Use c_t for scaling backward probabilities as well
- To get back true forward/backward probabilities

$$f^*_t(i) = (\prod^{t}_{t=1} c_t) f_t(i)$$

Reminders

- HW7 due this Sunday, 11:59pm
- Please have your name in the filename of your homework assignment and match the template