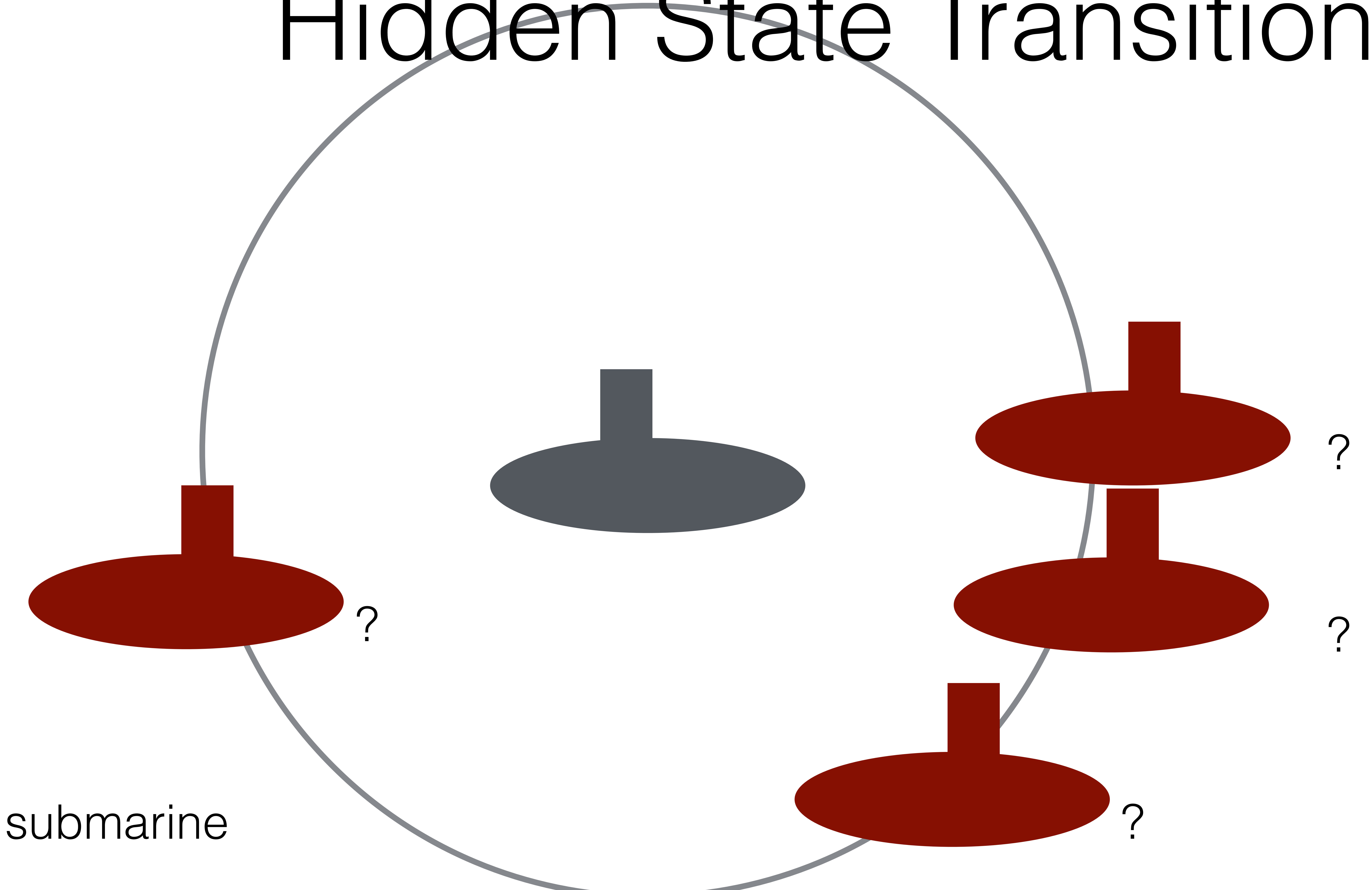
Hidden Markov Models

Machine Learning
CSx824/ECEx242
Bert Huang
Virginia Tech

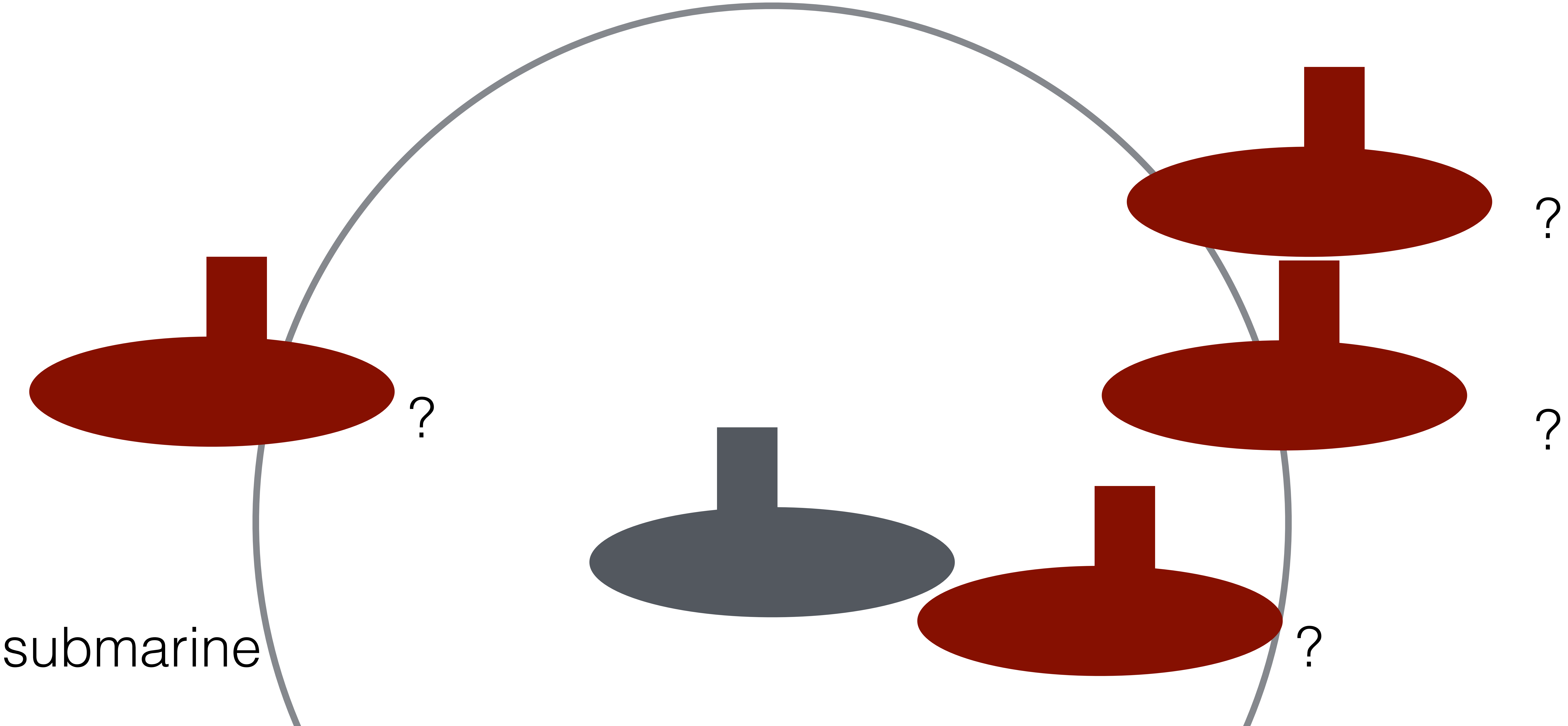
Outline

- Hidden Markov models (HMMs)
- Forward-backward for HMMs
- Baum-Welch learning (expectation maximization)

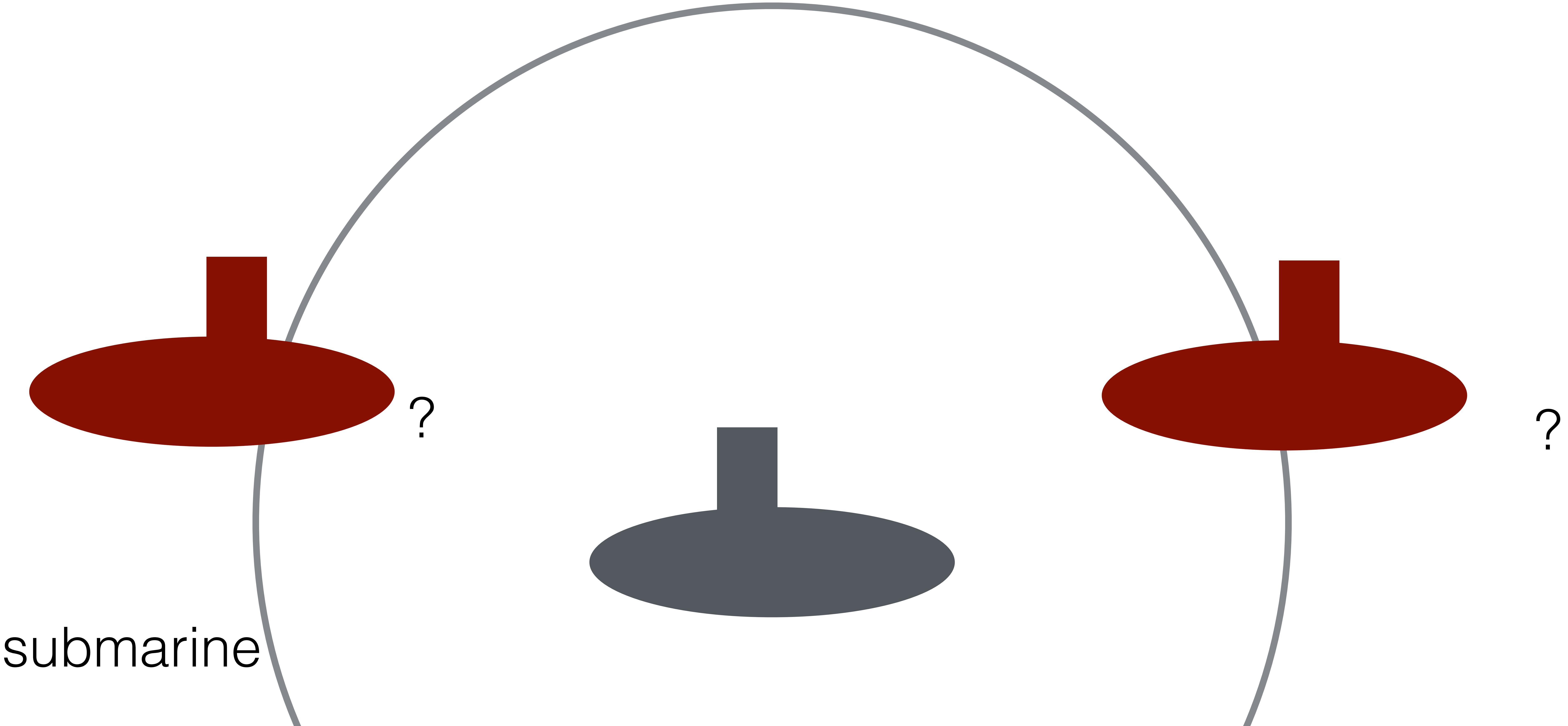
Hidden State Transitions



Hidden State Transitions



Hidden State Transitions

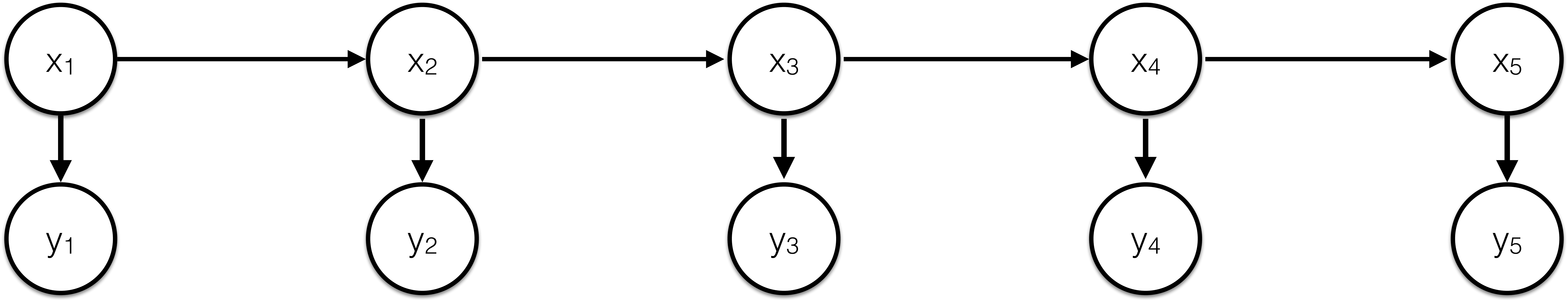


Hidden Markov Models

$p(y_t|x_t)$ observation probability SONAR noisiness

$p(x_t|x_{t-1})$ transition probability submarine locomotion

$$p(X, Y) = p(x_1) \prod_{t=1}^{T-1} p(x_{t+1}|x_t) \prod_{t'=1}^{T} p(y_{t'}|x_{t'})$$



Hidden State Inference

$$p(X|Y) \qquad p(x_t|Y)$$

$$\alpha_t(x_t) = p(x_t, y_1, \ldots, y_t) \qquad \beta_t(x_t) = p(y_{t+1}, \ldots, y_T|x_t)$$

$$\alpha_t(x_t)\beta_t(x_t) = p(x_t, y_1, \ldots, y_t)p(y_{t+1}, \ldots, y_T|x_t) \; = p(x_t, Y) \; \propto p(x_t|Y)$$

normalize to get conditional probability

note: not the same as $p(x_1, \ldots, x_T, Y)$

Forward Inference

$$\alpha_t(x_t) = p(x_t, y_1, \ldots, y_t)$$

$$p(x_1, y_1) = p(x_1)p(y_1|x_1) = \alpha_1(x_1)$$

$$p(x_2, y_1, y_2) = \sum_{x_1} p(x_1, y_1)p(x_2|x_1)p(y_2|x_2) = \alpha_2(x_2) = \sum_{x_1} \alpha_1(x_1)p(x_2|x_1)p(y_2|x_2)$$

$$p(x_{t+1}, y_1, \ldots, y_{t+1}) = \alpha_{t+1}(x_{t+1}) = \sum_{x_t} \alpha_t(x_t)p(x_{t+1}|x_t)p(y_{t+1}|x_{t+1})$$

Backward Inference

$$\beta_t(x_t) = p(y_{t+1}, \dots, y_T | x_t)$$

$$p(\{\}|x_T) = 1 = \beta_T(x_T)$$

$$\beta_{t-1}(x_{t-1}) = p(y_t, \dots, y_T | x_{t-1}) = \sum_{x_t} p(x_t | x_{t-1}) p(y_t, y_{t+1}, \dots, y_T | x_t)$$

$$= \sum_{x_t} p(x_t | x_{t-1}) p(y_t | x_t) p(y_{t+1}, \dots, y_T | x_t)$$

$$= \sum_{x_t} p(x_t | x_{t-1}) p(y_t | x_t) \beta_t(x_t)$$

Backward Inference

$$\beta_t(x_t) = p(y_{t+1}, \dots, y_T | x_t)$$

$$p(\{\}|x_T) = 1 = \beta_T(x_T)$$

$$\beta_{t-1}(x_{t-1}) = p(y_t, \dots, y_T | x_{t-1}) = \sum_{x_t} p(x_t | x_{t-1}) p(y_t | x_t) \beta_t(x_t)$$

Fusing the Messages

$$\alpha_t(x_t) = p(x_t, y_1, \ldots, y_t) \qquad \beta_t(x_t) = p(y_{t+1}, \ldots, y_T | x_t)$$

$$\alpha_t(x_t)\beta_t(x_t) = p(x_t, y_1, \ldots, y_t)p(y_{t+1}, \ldots, y_T | x_t) \;\; = p(x_t, Y) \;\; \propto p(x_t | Y)$$

$$\begin{aligned} p(x_t, x_{t+1} | Y) &= \frac{p(x_t, x_{t+1}, y_1, \ldots, y_t, y_{t+1}, y_{t+2}, \ldots, y_T)}{p(Y)} \\ &= \frac{p(x_t, y_1, \ldots, y_t)p(x_{t+1}|x_t)p(y_{t+2}, \ldots, y_T | x_{t+1})p(y_{t+1}|x_{t+1})}{\sum_{x_T} p(x_t, Y)} \\ &= \frac{\alpha_t(x_t)p(x_{t+1}|x_t)\beta_{t+1}(x_{t+1})p(y_{t+1}|x_{t+1})}{\sum_{x_T} \alpha_T(x_T)} \end{aligned}$$

Forward-Backward Inference

$$\alpha_1(x_1) = p(x_1)p(y_1|x_1)$$

$$\alpha_{t+1}(x_{t+1}) = \sum_{x_t} \alpha_t(x_t)p(x_{t+1}|x_t)p(y_{t+1}|x_{t+1})$$

$$\beta_T(x_T) = 1$$

$$\beta_{t-1}(x_{t-1}) = \sum_{x_t} p(x_t|x_{t-1})p(y_t|x_t)\beta_t(x_t)$$

$$p(x_t, Y) = \alpha_t(x_t)\beta_t(x_t)$$

$$p(x_t|Y) = \frac{\alpha_t(x_t)\beta_t(x_t)}{\sum_{x'_t} \alpha_t(x'_t)\beta_t(x'_t)}$$

$$p(x_t, x_{t+1}|Y) \quad = \frac{\alpha_t(x_t)p(x_{t+1}|x_t)\beta_{t+1}(x_{t+1})p(y_{t+1}|x_{t+1})}{\sum_{x_T} \alpha_T(x_T)}$$

Normalization

To avoid underflow, re-normalize at each time step

$$\tilde{\alpha}_t(x_t) = \frac{\alpha_t(x_t)}{\sum_{x'_t} \alpha_t(x'_t)} \qquad \tilde{\beta}_t(x_t) = \frac{\beta_t(x_t)}{\sum_{x'_t} \beta_t(x'_t)}$$

Exercise: why is this okay?

Learning

- Parameterize and learn

$p(x_{t+1}|x_t)$

conditional probability table
transition matrix

$p(y_t|x_t)$

observation model
emission model

- If fully observed, super easy!
- If ***x*** is hidden (most cases) treat as latent variable
    - E.g., expectation maximization

Baum-Welch Algorithm

EM using forward-backward inference as E-step

Baum-Welch Details

Compute $p(x_t|Y)$ and $p(x_t, x_{t+1}|Y)$ using forward-backward

Maximize weighted (expected) log-likelihood

$$p(x_1) \leftarrow \frac{1}{T}\sum_{t=1}^{T} p(x_t|Y) \quad \text{or} \quad p(x_1|Y)$$

e.g., Gaussian

$$\mu_x \leftarrow \frac{\sum_{t=1}^{T} p(x_t = x|Y) y_t}{\sum_{t'=1}^{T} p(x_t = x|Y)}$$

$$p(x_{t'+1} = i | x_{t'} = j) \leftarrow \frac{\sum_{t=1}^{T-1} p(x_{t+1} = i, x_t = j|Y)}{\sum_{t=1}^{T-1} p(x_t = j|Y)}$$

$$p(y|x) \leftarrow \frac{\sum_{t=1}^{T} p(x_t = x|Y) I(y_t = y)}{\sum_{t'=1}^{T} p(x_t = x|Y)}$$

e.g., multinomial

Summary

- HMMs represent hidden states
  - Transitions between adjacent states
  - Observation based on states
- Forward-backward inference to incorporate all evidence
- Expectation maximization to train parameters (Baum-Welch)
  - Treat states as latent variables