HMMs

States: E_1, E_2, ... E_N

Initial state probabilities: $\pi(i)$

Transition probabilities: a_{ij}

Alphabet, Σ

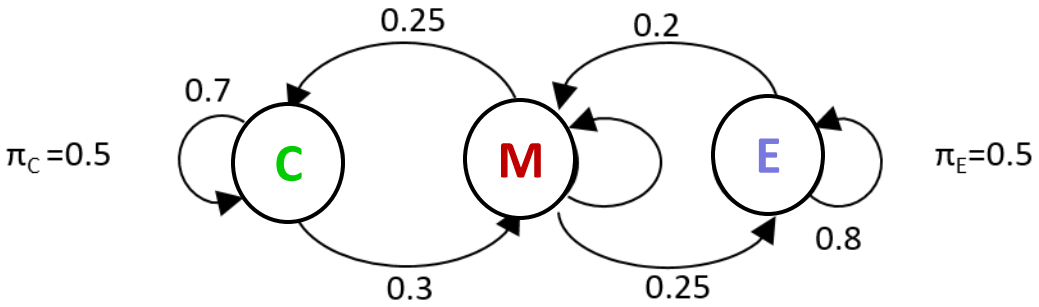
Emission probabilities: e_i

The parameters of the HMM $\lambda=(a_{ij}, e_i(\sigma), \pi)$ are "learned" from known examples ("labeled data").

An HMM is a *generative* model: we say

"the model emitted sequence $O = O_1 O_2 O_3 \ldots O_T$ via state path $Q = q_1 q_2 q_3 \ldots q_T$ "

A three state transmembrane HMM:



e_C(H)	0.3
e_C(L)	0.7

e_M(H)	0.9
e_M(L)	0.1

e_E(H)	0.2
e_E(L)	0.8

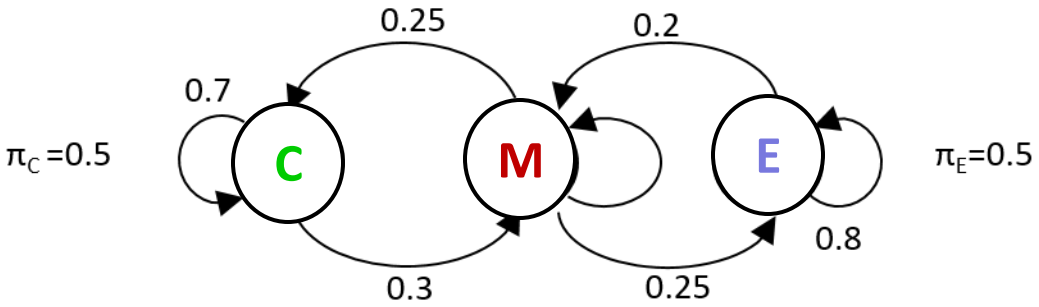
An HMM generates *labeled* sequences:

```
LLLHLHLLHLLLHHHLLHHHHLHHHLLHLLHLL...
CCCCCCCCCCCCMMMMMMMMMMMMMMEEEEEEEE...

                    LLLHLHHHHHHHLLHLLLLLHLHHHLLHLLHL...
                    CCCCCMMMMMMMEEEEEEEEMMMMMCCCCCCCC...
          LHLLLHLLHLHLHHHHHHHLHLHLLHHLLHHHHHLHLLLLHL...
          EEEEEEEEEEMMMMMMMCCCCCCCCCCMMMMMMMEEEEEEE...

 LLLHLHLLHLHHHLLHHHHLHHHLLHLLHLLLLLLLLL...
 CCCCCCCCMMMMMMMMMMMMMMMMMMMEEEEEEEEE...
```

A three state transmembrane HMM:



$e_C(H)$	0.3
$e_C(L)$	0.7

$e_M(H)$	0.9
$e_M(L)$	0.1

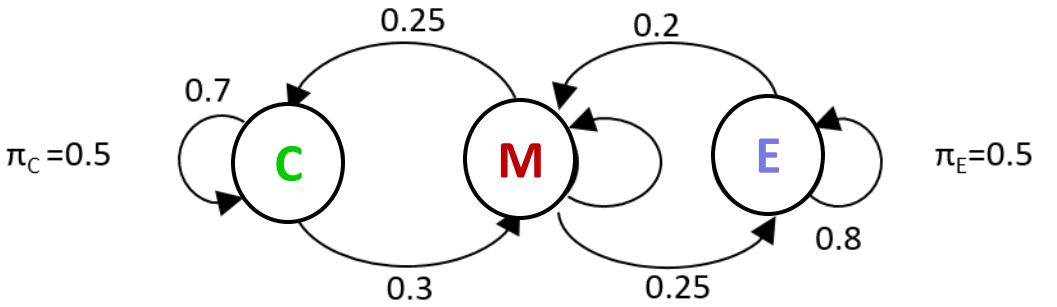
$e_E(H)$	0.2
$e_E(L)$	0.8

What is the probability that this model emitted LHHHL via path CMMME?

What is $P(O, Q|\lambda)$, where O = LHHHL and Q = CMMME?

$$P(O, Q|\lambda) = \pi_{q_1} \cdot e_{q_1}(O_1) \prod_{i=2}^{T} a_{q_{i-1}q_i} e_{q_i}(O_i)$$

A three state transmembrane HMM:



$e_C(H)$	0.3
$e_C(L)$	0.7

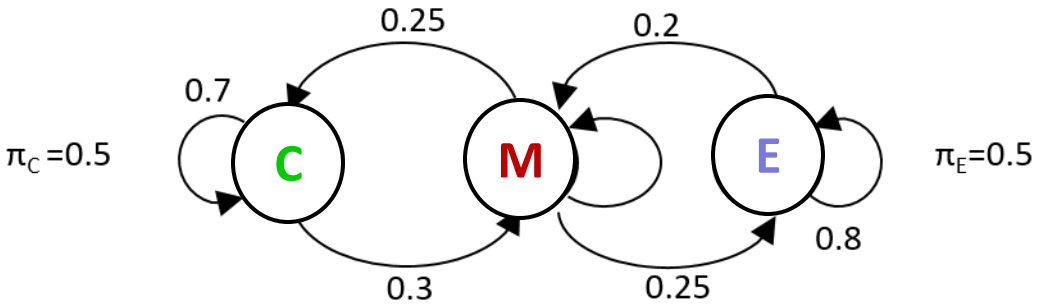
$e_M(H)$	0.9
$e_M(L)$	0.1

$e_E(H)$	0.2
$e_E(L)$	0.8

Given unlabeled data, and an HMM

LLLHLHLLHLLLHHHLLHHHHLHHHLLHLLHLL...

A three state transmembrane HMM:



$e_C(H)$	0.3
$e_C(L)$	0.7

$e_M(H)$	0.9
$e_M(L)$	0.1

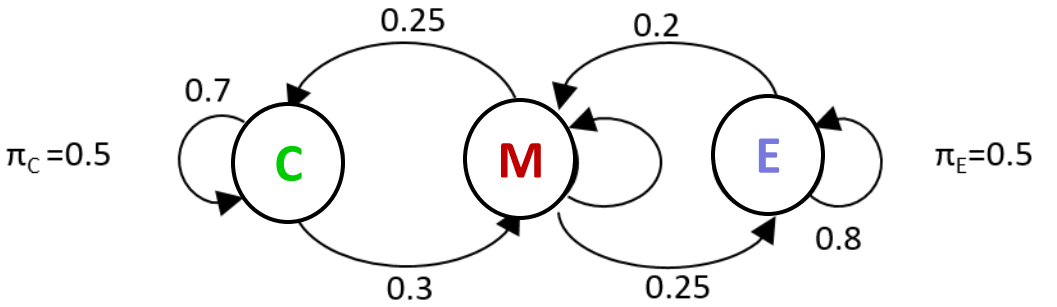
$e_E(H)$	0.2
$e_E(L)$	0.8

Given unlabeled data, and an HMM

What is P(O|HMM), the probability of a given sequence?

LLLHLHLLHLLLHHHLLHHHHLHHHLLHLLHLL...

A three state transmembrane HMM:



$e_C(H)$	0.3
$e_C(L)$	0.7

$e_M(H)$	0.9
$e_M(L)$	0.1

$e_E(H)$	0.2
$e_E(L)$	0.8

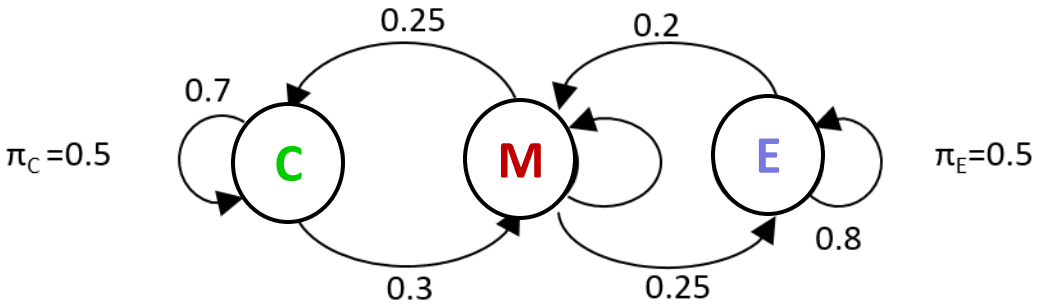
Given unlabeled data, and an HMM

What is P(O|HMM), the probability of a given sequence?

What is the state path?

```
LLLHLHLLHLLLHHHLLHHHHLHHHLLHLLHLL...
CCCCCCCCCCCCMMMMMMMMMMMMMMEEEEEEEE...
```

A three state transmembrane HMM:



$e_C(H)$	0.3
$e_C(L)$	0.7

$e_M(H)$	0.9
$e_M(L)$	0.1

$e_E(H)$	0.2
$e_E(L)$	0.8

Given unlabeled data, and an HMM

What is P(O|HMM), the probability of a given sequence?

What is the state path?

What state emitted the symbol O_t

```
LLLHLHLLHLLLHHHLLHHHHLHHHLLHLLHLL...
          M
```

Recognition problems

- What is the probability of a given sequence?

  Example: given HHLHH, is it a TM sequence or not?

- Given a sequence of symbols, what is the "true" sequence of states?

  Example: given HHHLLHL…, where is the TM region?

- What state emitted the symbol O_t?

  Example: is the isoleucine at position 32 localized to the membrane?

Recognition problems

- What is the probability of a given sequence, O?

  Forward algorithm

- Given a sequence O, what is the "true" sequence of states?

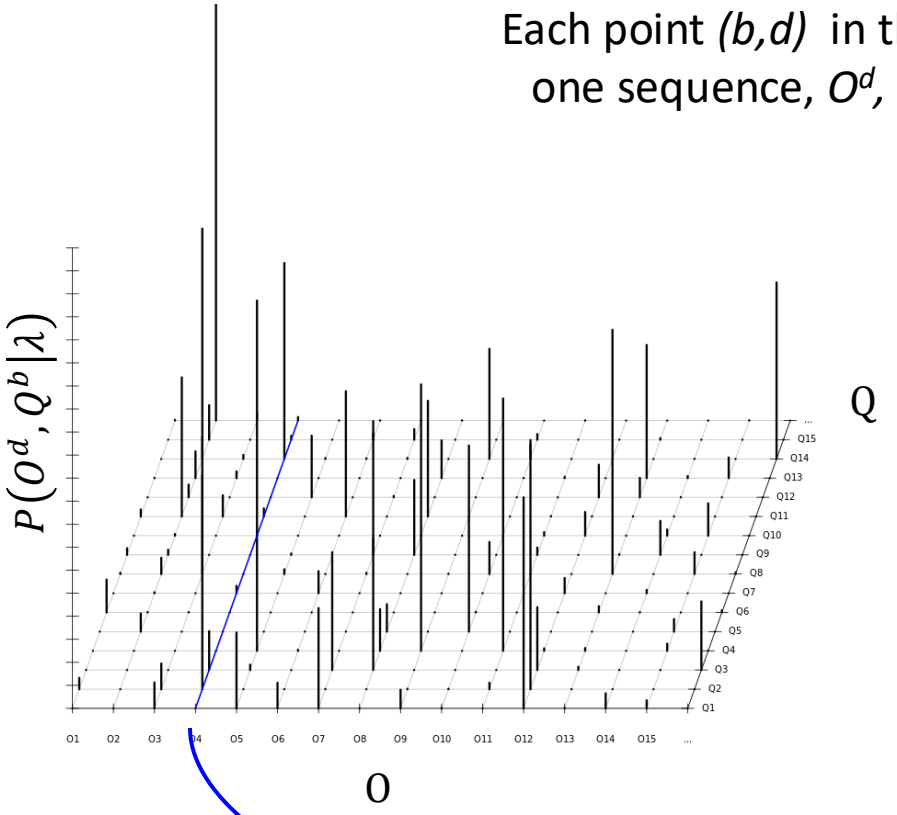
  Viterbi decoding: Viterbi algorithm

  Posterior decoding: Forward and Backward algorithms

- What state emitted the symbol O_t?

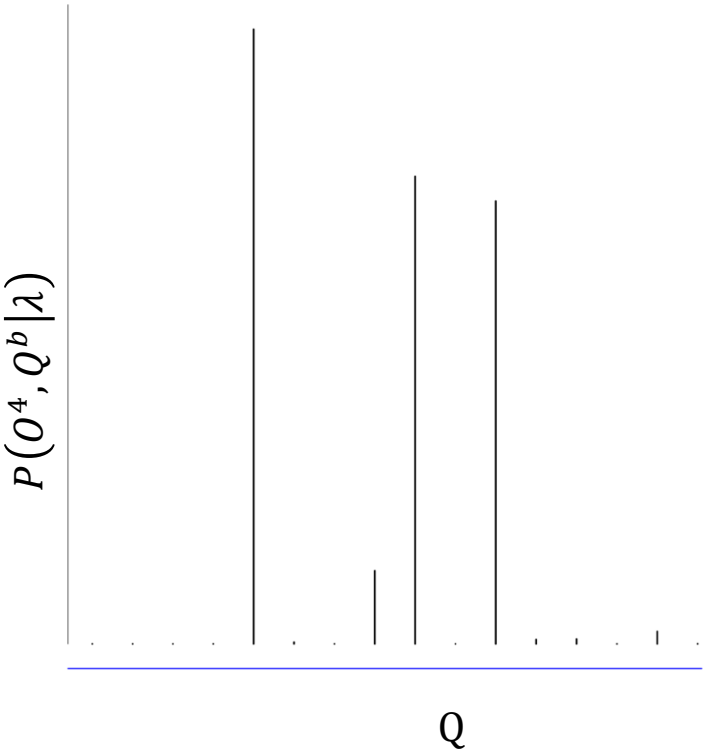
  Posterior decoding: Forward and Backward algorithms

Each point *(b,d)* in the plane corresponds to one sequence, O^d, and one state path, Q^b



The probability of emitting *some* sequence via *some* state path is 1:

$$\sum_b \sum_d P(O^d, Q^b|\lambda) = 1$$



This plane corresponds to all ways to emit sequence, O^d . Each point *b* on the horizontal axis corresponds to one state path, Q^b

Recognition problems

- What is the probability of a given sequence, O?

  Forward algorithm

- Given a sequence O, what is the "true" sequence of states?

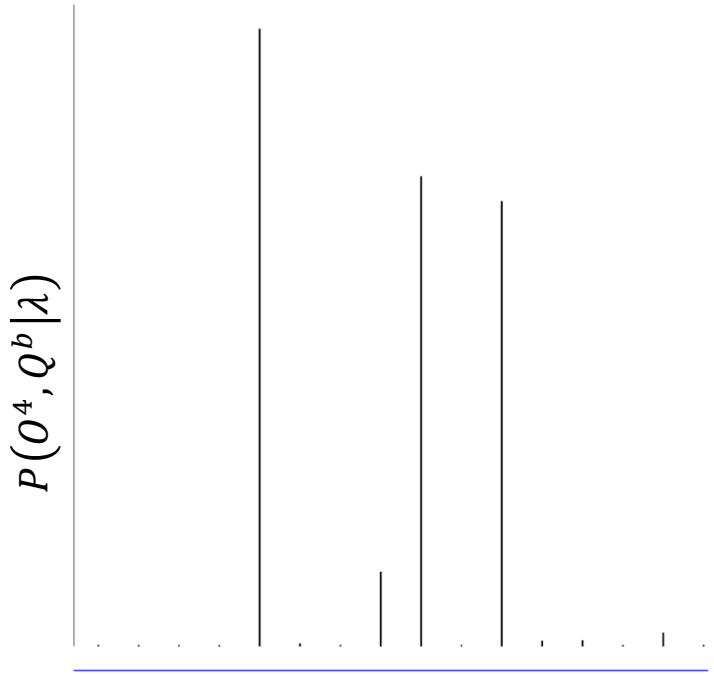
  Viterbi decoding: Viterbi algorithm

  Posterior decoding: Forward and Backward algorithms

- What state emitted the symbol O_t?

  Posterior decoding: Forward and Backward algorithms

HMM Dynamic Programming algorithms for recognition problems



The *Forward* algorithm calculates the probability of emiiting O^4 by summing over all possible paths

$$P(O^4) = \sum_j P(O^4, Q^b|\lambda)$$

Recognition problems

- What is the probability of a given sequence, O?

  Forward algorithm

- Given a sequence O, what is the "true" sequence of states?

  Viterbi decoding: Viterbi algorithm

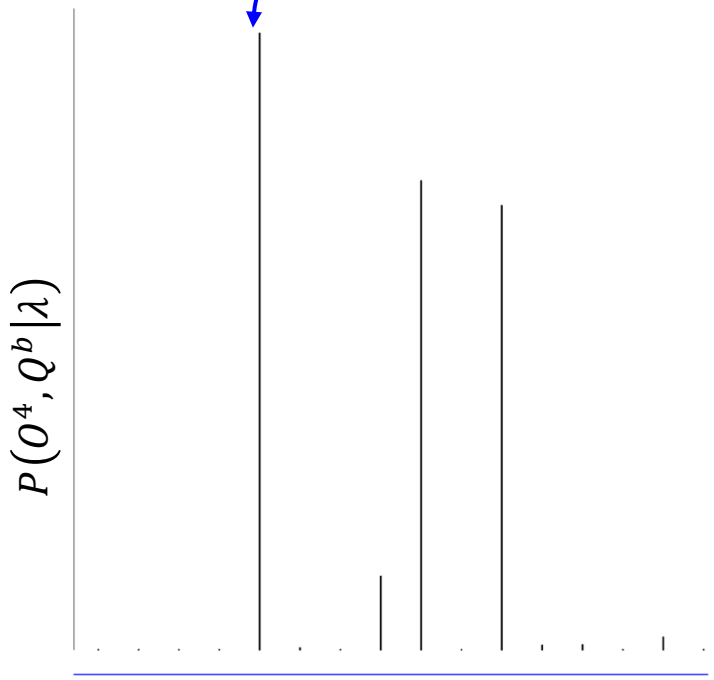
  Posterior decoding: Forward and Backward algorithms

- What state emitted the symbol O_t?

  Posterior decoding: Forward and Backward algorithms

HMM Dynamic Programming algorithms for recognition problems

The *Viterbi* algorithm finds the path that maximizes $P(O^4, Q^b|\lambda)$



The *Forward* algorithm calculates the probability of emiiting O^4 by summing over all possible paths

$$P(O^4) = \sum_j P(O^4, Q^b|\lambda)$$

Recognition problems

- What is the probability of a given sequence, O?

  Forward algorithm

- Given a sequence O, what is the "true" sequence of states?

  Viterbi decoding: Viterbi algorithm

  Posterior decoding: Forward and Backward algorithms

- What state emitted the symbol O_t?

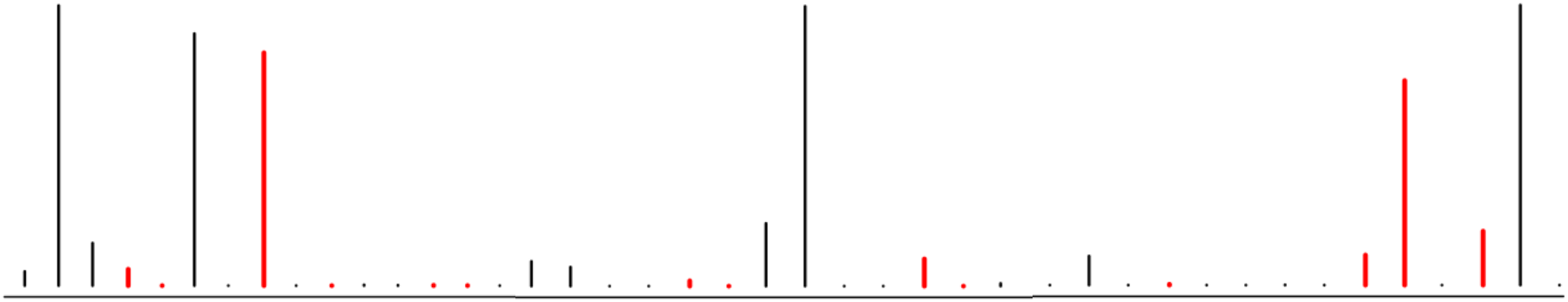
  Posterior decoding: Forward and Backward algorithms

What state emitted the symbol O_t?

$$P(O, q_t = E_i|\lambda)$$

$Q = q_1 q_2 \ldots q_t = E_i \ldots q_T$

$Q = q_1 q_2 \ldots q_t \neq E_i \ldots q_T$



Sum over all paths that pass through E_i at t