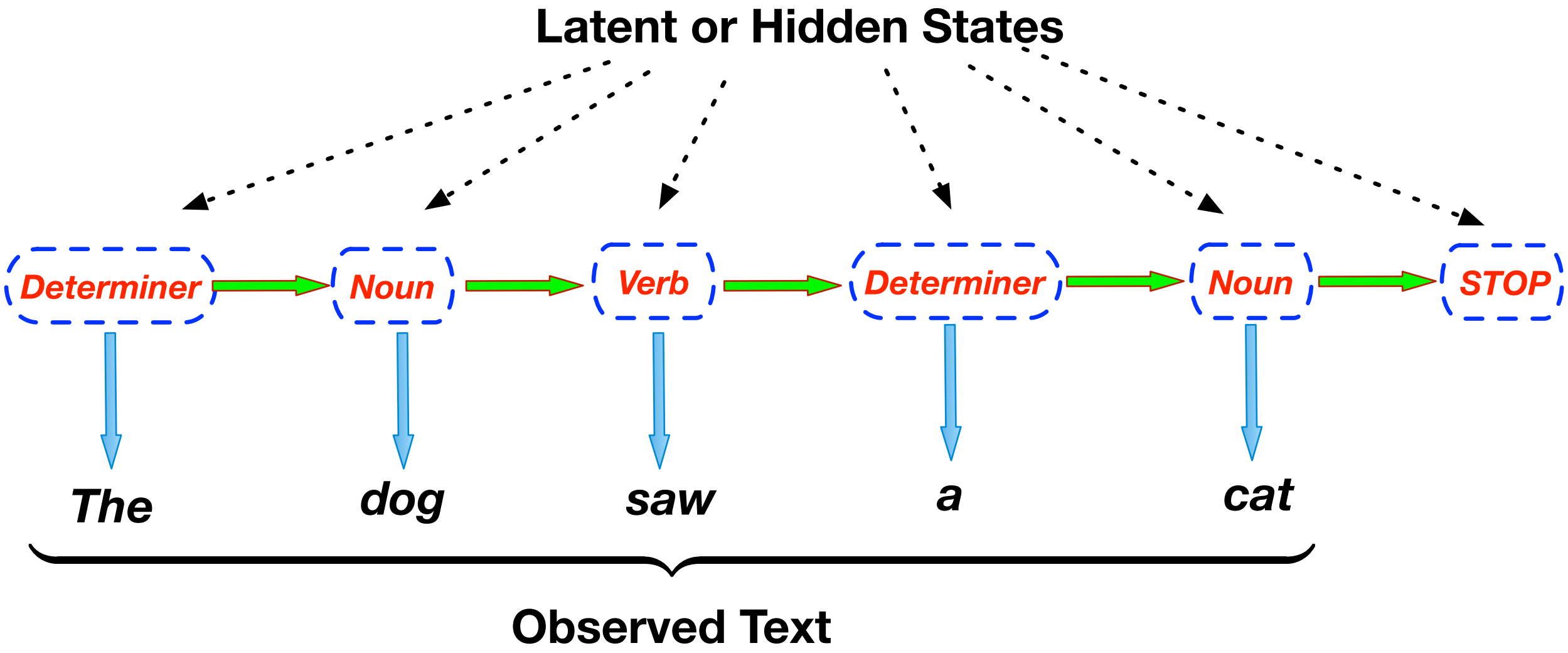
Special Topics in Natural Language Processing

CS698O

Ashutosh Modi
CSE Department, IIT Kanpur

Lecture 11: Sequence Prediction 2
Jan 29, 2020

Hidden Markov Models (HMM)



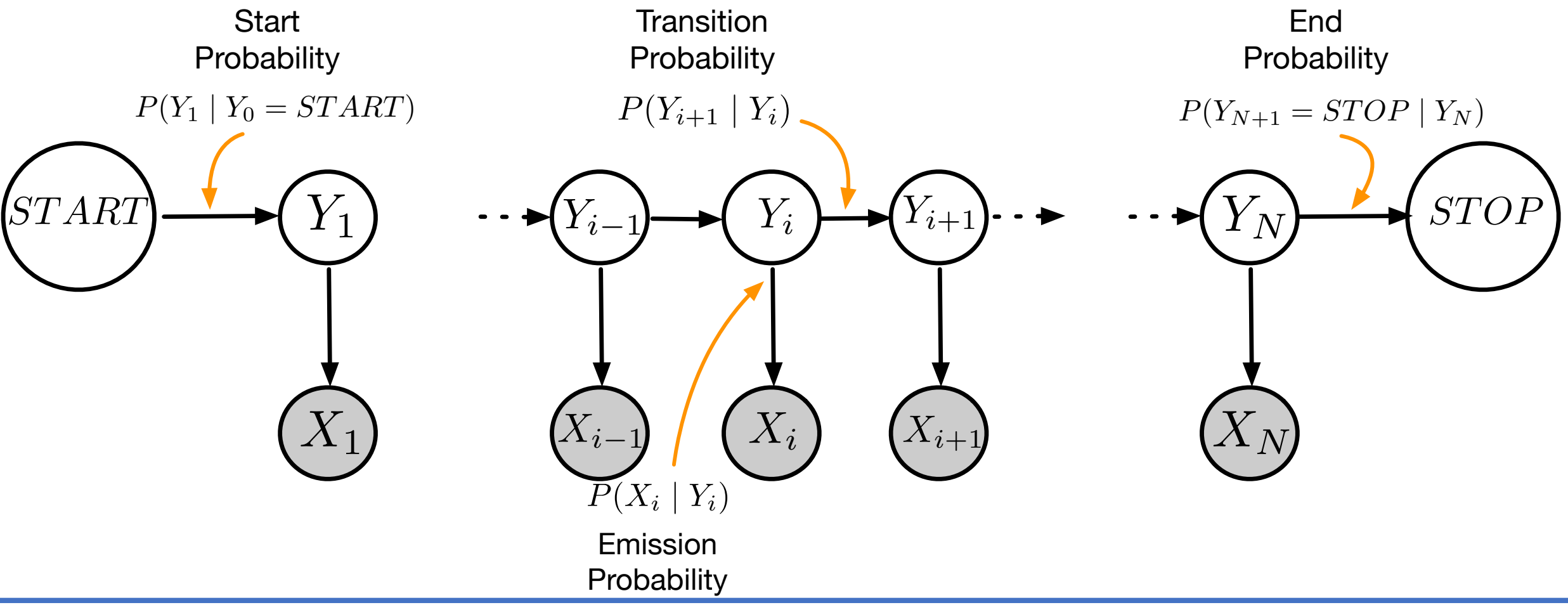
HMM Setting

Set of States (Λ)

$$\{c_0 = START, c_1, c_2, ..., c_K, c_{K+1} = STOP\}$$

Set of Observations

$$\{w_1, w_2, ...w_J\}$$



HMM Joint Distribution

- We have a generative model and are interested in the joint distribution

$$P(X, Y) = P\left(X_1 = x_1, \ldots, X_N = x_N, Y_1 = y_1, \ldots, Y_N = y_N\right)$$

$$P\left(X_1 = x_1, \ldots, X_N = x_N, Y_1 = y_1, \ldots, Y_N = y_N\right) =$$

$$P\left(y_1 | START\right) \times \left(\prod_{i=1}^{N-1} P\left(y_{i+1} | y_i\right)\right) \times \left(\prod_{i=1}^{N} P\left(x_i | y_i\right)\right) \times P\left(STOP | y_N\right)$$

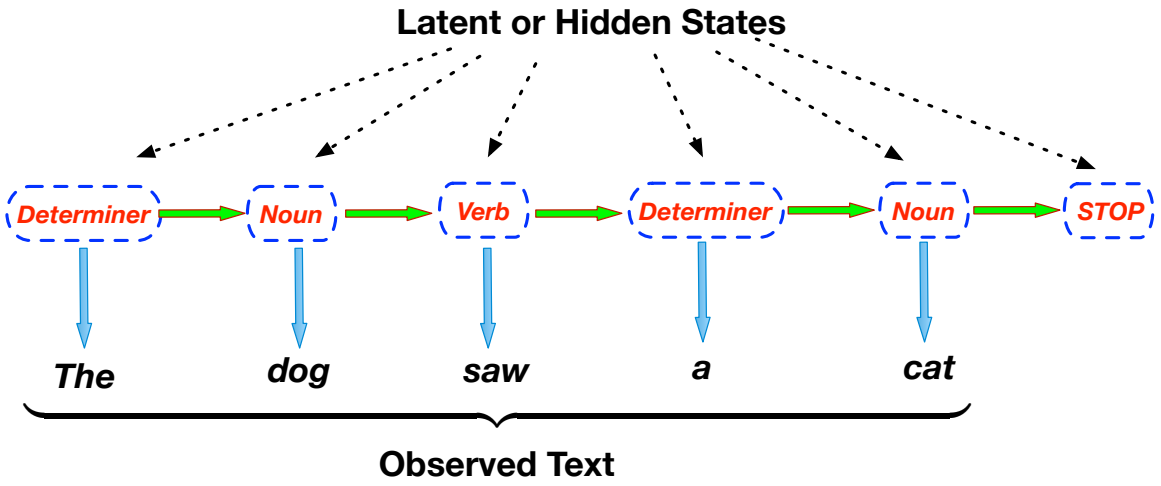
Initial / Start Probability

Transition Probability

Emission Probability

End / Final Probability

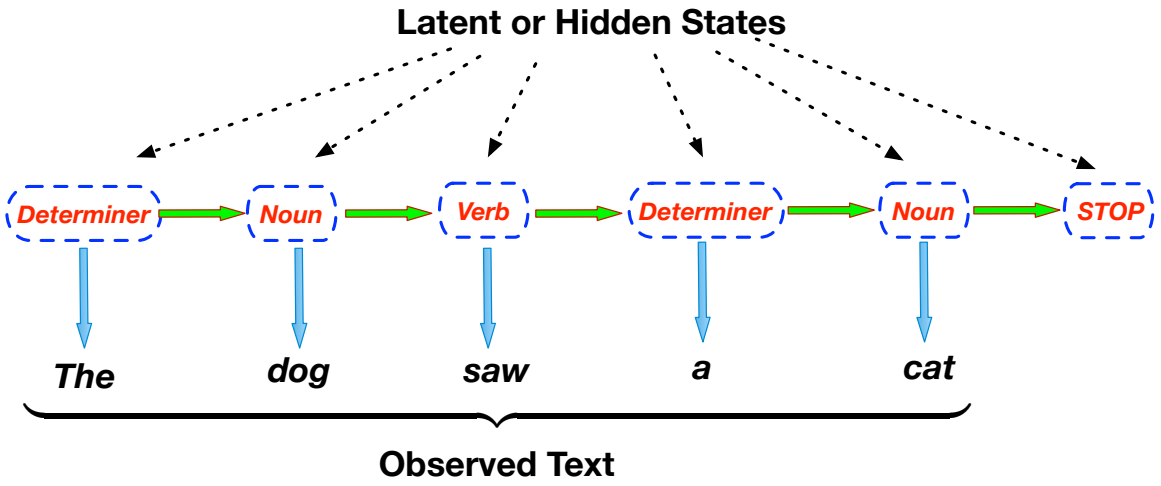
Hidden Markov Models (HMM)



Three Questions:

1. ***Learning Problem***: How do we estimate the parameters of distributions?
2. ***Decoding Problem***: Given the observed text what is the hidden POS sequence that best explains the observation?
3. ***Likelihood Problem***: Given the parameters of the distributions what is the probability of the observed sequence?

Hidden Markov Models (HMM)



Three Questions:

1. ***Learning Problem***: How do we estimate the parameters of distributions?
2. ***Decoding Problem***: Given the observed text what is the hidden POS sequence that best explains the observation?
3. ***Likelihood Problem***: Given the parameters of the distributions what is the probability of the observed sequence?

Learning Problem

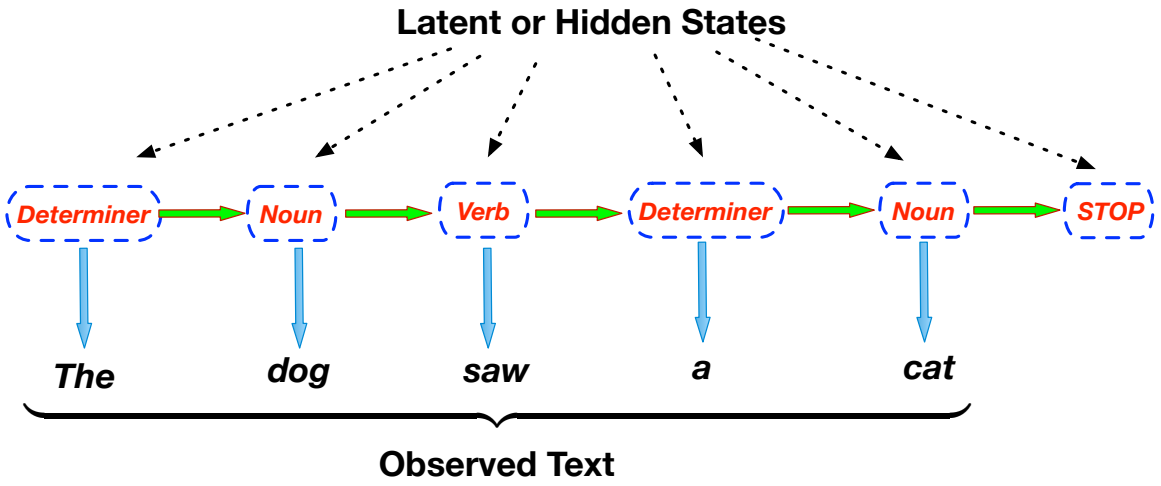
$$P_{\text{init}}\ (y_1 = c_k | START) = \frac{Count(y_1 = c_k)}{\left(\sum_{l=1}^{K} Count(y_1 = c_l)\right) = M}$$

$$P_{\text{trans}}\ (y_{i+1} = c_k | y_i = c_l) = \frac{Count(c_k, c_l)}{Count(c_l)}$$

$$P_{\text{emiss}}\ (x_i = w_j | y_i = c_k) = \frac{Count(w_j, c_k)}{Count(c_k)}$$

$$P_{\text{final}}\ (STOP | y_N = c_k) = \frac{Count(y_N = c_k)}{\left(\sum_{l=1}^{K} Count(y_N = c_l)\right) = M}$$

Hidden Markov Models (HMM)



Three Questions:

1. ***Learning Problem***: How do we estimate the parameters of distributions?
2. ***Decoding Problem***: Given the observed text what is the hidden POS sequence that best explains the observation?
3. ***Likelihood Problem***: Given the parameters of the distributions what is the probability of the observed sequence?

Decoding Problem

- Two ways possible

Decoding Problem

- Two ways possible
  - **Posterior Decoding**: Maximize probability of each state

$$y_i^* = \arg\max_{y_i \in \Lambda} P\left(Y_i = y_i | X_1 = x_1, \ldots, X_N = x_N\right)$$

Decoding Problem

- Two ways possible
  - **Posterior Decoding**: Maximize probability of each state

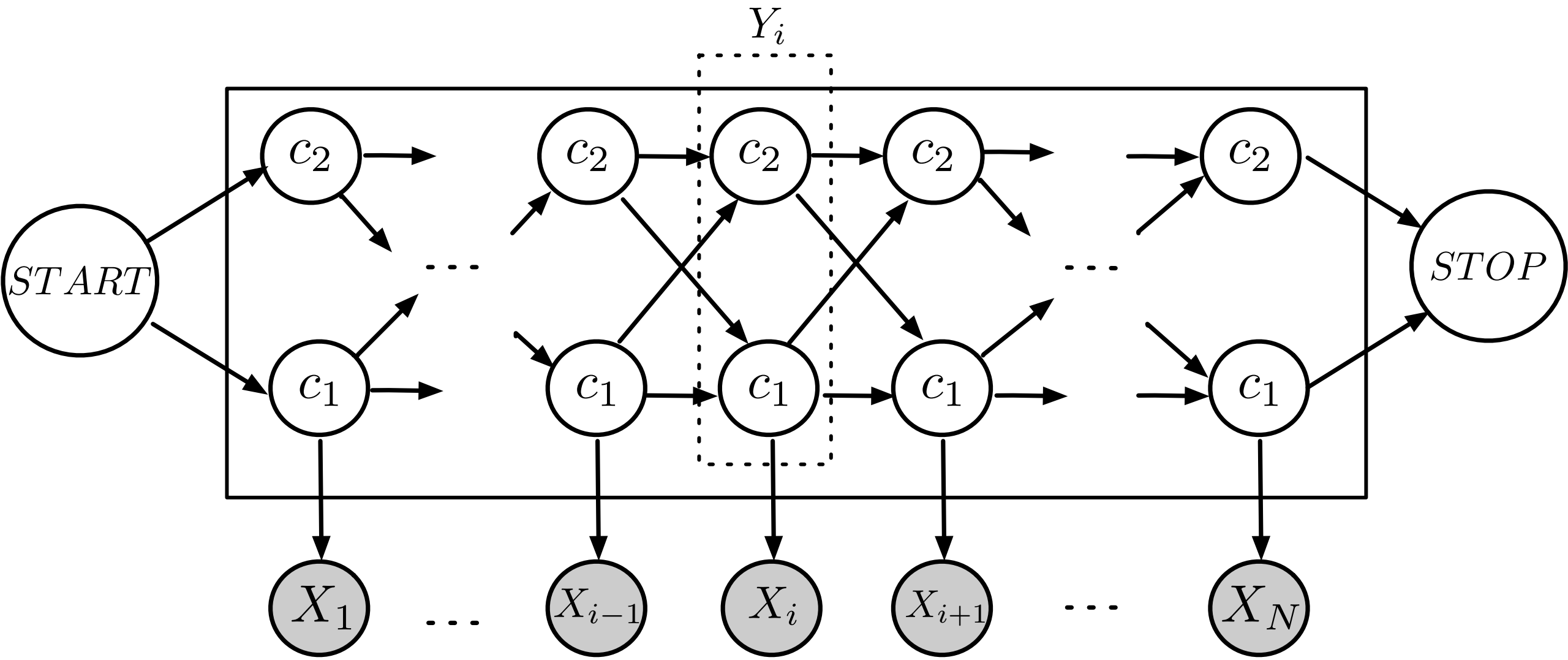
$$y_i^* = \arg\max_{y_i \in \Lambda} P\left(Y_i = y_i | X_1 = x_1, \ldots, X_N = x_N\right)$$

  - **Viterbi Decoding**: Maximize the probability of the entire sequence

$$y^* = \arg\max_{y = y_1 \ldots y_N} P\left(Y_1 = y_1, \ldots, Y_N = y_N | X_1 = x_1, \ldots, X_N = x_N\right)$$

$$= \arg\max_{y = y_1 \ldots y_N} P\left(Y_1 = y_1, \ldots, Y_N = y_N, X_1 = x_1, \ldots, X_N = x_N\right)$$

Trellis Diagram



Decoding Problem

- Decoding requires the following calculation

$$P(Y \mid X) = \frac{P(X, Y)}{P(X)}$$

$$P(Y \mid X) = \frac{P(X, Y)}{\sum_{Y} P(X, Y)}$$

Decoding Problem

- Decoding requires the following calculation

$$P(Y \mid X) = \frac{P(X, Y)}{P(X)}$$

$$P(Y \mid X) = \frac{P(X, Y)}{\sum_{Y} P(X, Y)}$$

But what is the sum in the denominator over?

Over all possible sequences $Y = y_1, \ldots, y_N$

Decoding Problem

$$\text{In general, number of sequences} = K^N$$

Decoding Problem

$$\text{In general, number of sequences} = K^N$$

EXPONENTIAL IN LENGTH

Decoding Problem

$$\text{In general, number of sequences} = K^N$$

EXPONENTIAL IN LENGTH

What do we do?

Decoding Problem

$$\text{In general, number of sequences} = K^N$$

EXPONENTIAL IN LENGTH

What do we do?

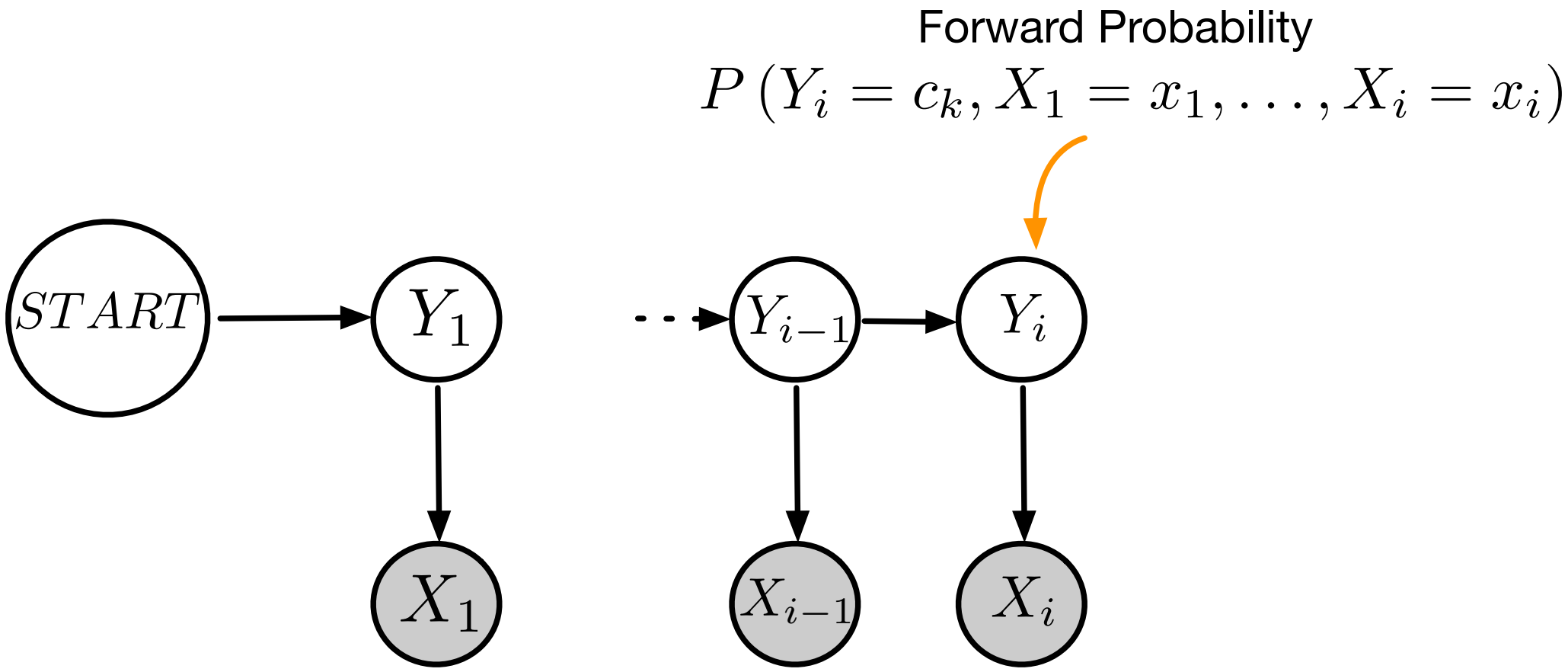
DYNAMIC PROGRAMMING

Forward Probability

$$\text{forward}\,(i, c_k) := P\left(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i\right)$$

Forward Probability

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$



Forward Probability

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\,(Y_i = c_k \mid Y_{i-1} = c_l) \times \text{forward}(i-1, c_l) \right) \times P\,(X_i \mid Y_i = c_k)$$

Forward Probability

$$\text{forward}\,(i, c_k) := P\left(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i\right)$$

Forward Probability

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i) = P\,(Y_i = c_k, X_{1:i})$$

Forward Probability

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i) = P\,(Y_i = c_k, X_{1:i})$$

$$P\,(Y_i = c_k, X_{1:i})$$

$$= \sum_{Y_{i-1}} P\,(Y_i = c_k, Y_{i-1}, X_{1:i})$$

Marginalization

Forward Probability

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i) = P\,(Y_i = c_k, X_{1:i})$$

$$P\,(Y_i = c_k, X_{1:i})$$

$$= \sum_{Y_{i-1}} P\,(Y_i = c_k, Y_{i-1}, X_{1:i})$$

Marginalization

$$= \sum_{Y_{i-1}} P\,(X_i \mid Y_i, Y_{i-1}, X_{1:i-1}) \times P\,(Y_i, Y_{i-1}, X_{1:i-1})$$

Bayes Rule

Forward Probability

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i) = P\,(Y_i = c_k, X_{1:i})$$

$$P\,(Y_i = c_k, X_{1:i})$$

$$= \sum_{Y_{i-1}} P\,(Y_i = c_k, Y_{i-1}, X_{1:i})$$

Marginalization

$$= \sum_{Y_{i-1}} P\,(X_i \mid Y_i, Y_{i-1}, X_{1:i-1}) \times P\,(Y_i, Y_{i-1}, X_{1:i-1})$$

Bayes Rule

$$= \sum_{Y_{i-1}} P\,(X_i \mid Y_i, Y_{i-1}, X_{1:i-1}) \times P\,(Y_i \mid Y_{i-1}, X_{1:i-1}) \times P\,(Y_{i-1}, X_{1:i-1})$$

Forward Probability

$$\text{forward}(i, c_k) := P(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$P(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i) = P(Y_i = c_k, X_{1:i})$$

$$P(Y_i = c_k, X_{1:i})$$

$$= \sum_{Y_{i-1}} P(Y_i = c_k, Y_{i-1}, X_{1:i})$$

Marginalization

$$= \sum_{Y_{i-1}} P(X_i \mid Y_i, Y_{i-1}, X_{1:i-1}) \times P(Y_i, Y_{i-1}, X_{1:i-1})$$

Bayes Rule

$$= \sum_{Y_{i-1}} P(X_i \mid Y_i, Y_{i-1}, X_{1:i-1}) \times P(Y_i \mid Y_{i-1}, X_{1:i-1}) \times P(Y_{i-1}, X_{1:i-1})$$

$$= \sum_{Y_{i-1}} P(X_i \mid Y_i) \times P(Y_i \mid Y_{i-1}) \times \text{forward}(i-1, y_{i-1})$$

Independence

Forward Probability

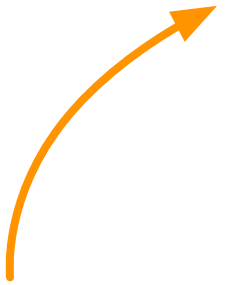
$$\text{forward}\,(i, c_k) := P\left(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i\right)$$

$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\left(Y_i = c_k \mid Y_{i-1} = c_l\right) \times \text{forward}(i-1, c_l)\right) \times P\left(X_i \mid Y_i = c_k\right)$$

Forward Probability

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\,(Y_i = c_k \mid Y_{i-1} = c_l) \times \text{forward}(i-1, c_l) \right) \times P\,(X_i \mid Y_i = c_k)$$



Transition Probability

Forward Probability

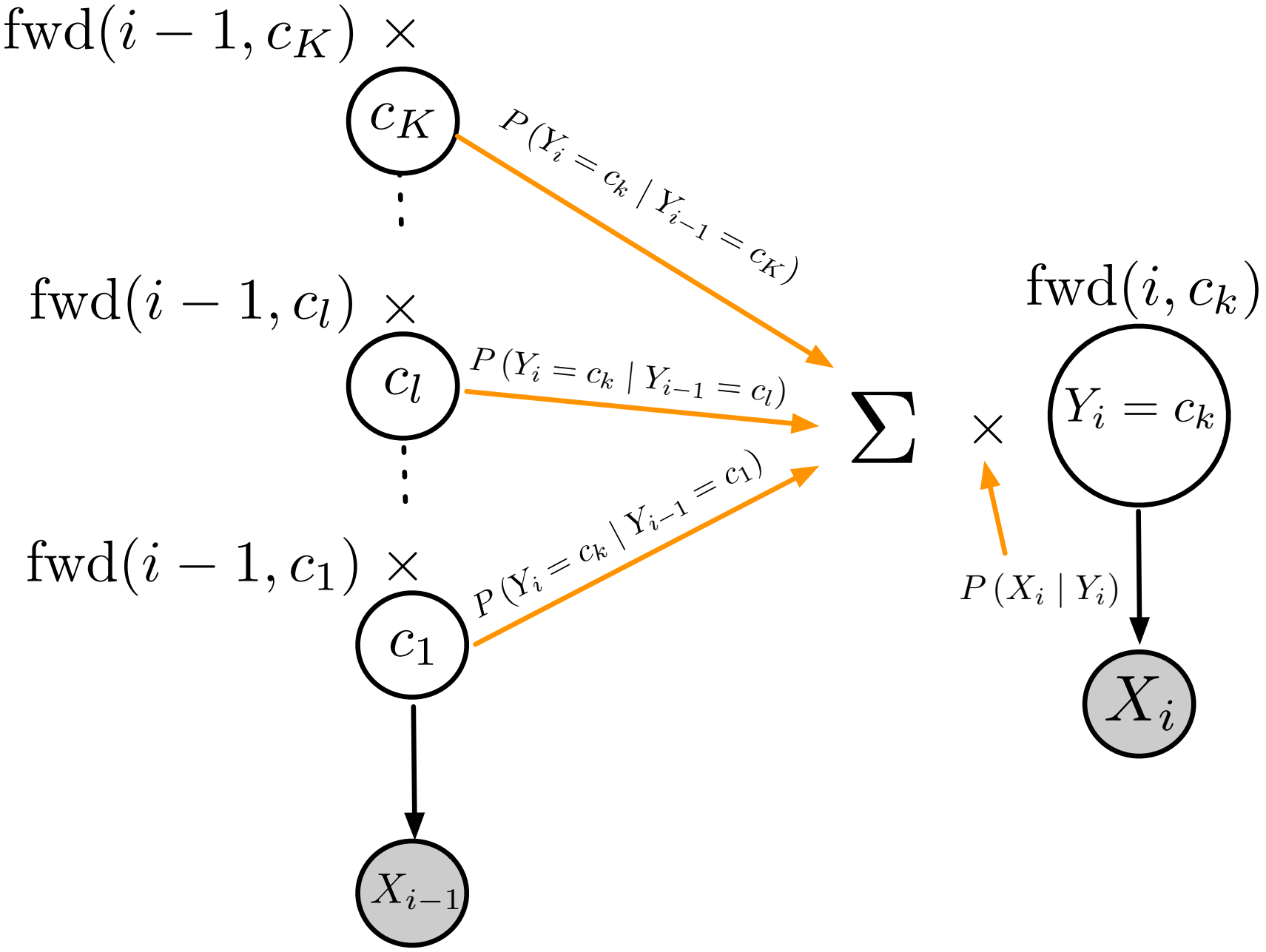
$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\,(Y_i = c_k \mid Y_{i-1} = c_l) \times \text{forward}(i-1, c_l) \right) \times P\,(X_i \mid Y_i = c_k)$$

Transition Probability

Emission Probability

Forward Probability



Forward Probability

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\,(Y_i = c_k \mid Y_{i-1} = c_l) \times \text{forward}(i-1, c_l) \right) \times P\,(X_i \mid Y_i = c_k)$$

$$\text{forward}(N+1, \mathit{STOP}) = ?$$

Forward Probability

$$\text{forward}\,(i, c_k) := P\left(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i\right)$$

$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\left(Y_i = c_k \mid Y_{i-1} = c_l\right) \times \text{forward}(i-1, c_l)\right) \times P\left(X_i \mid Y_i = c_k\right)$$

$$\text{forward}(N+1, STOP)$$

$$= P\left(Y_{N+1} = STOP, X_1 = x_1, \ldots, X_N = x_N, X_{N+1}\right)$$

Forward Probability

$$\text{forward}\,(i, c_k) := P\left(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i\right)$$

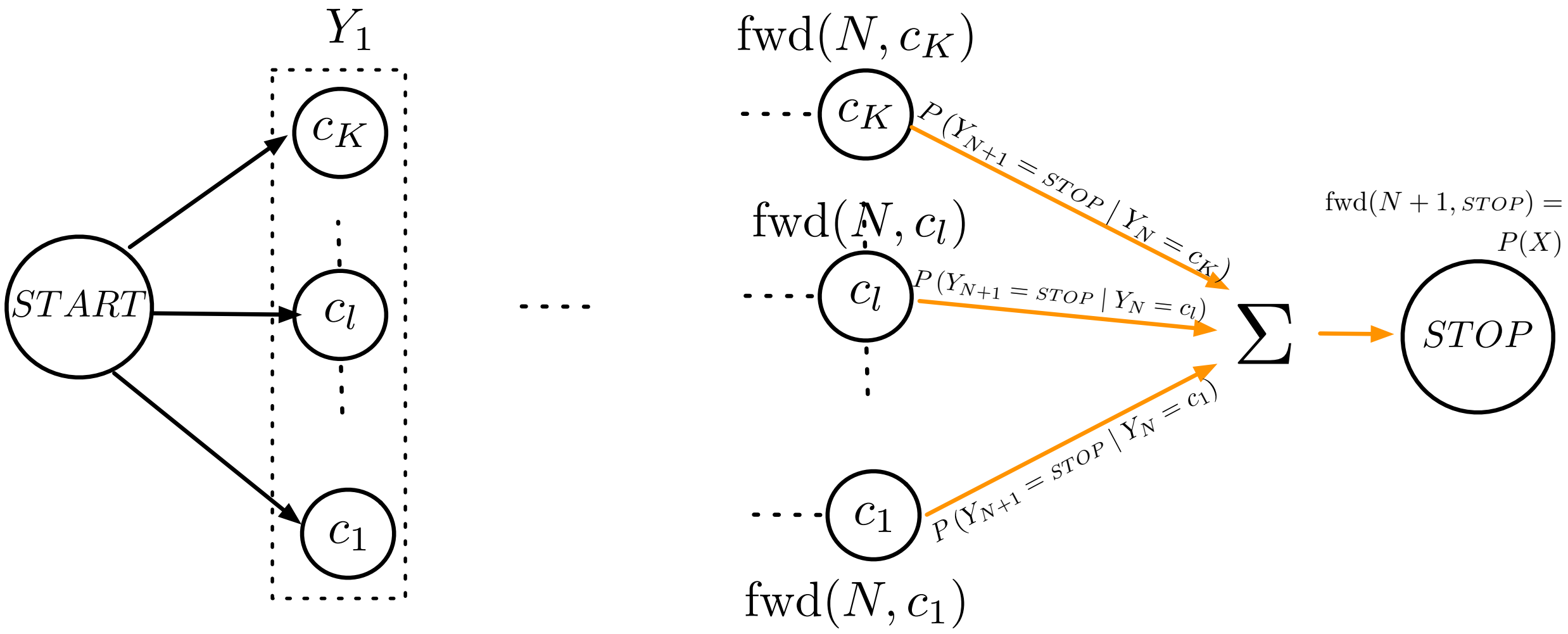
$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\left(Y_i = c_k \mid Y_{i-1} = c_l\right) \times \text{forward}(i-1, c_l)\right) \times P\left(X_i \mid Y_i = c_k\right)$$

$$\text{forward}(N+1, STOP)$$

$$= P\left(Y_{N+1} = STOP, X_1 = x_1, \ldots, X_N = x_N, X_{N+1}\right)$$

$$= P(X)$$

Forward Probability



Forward Probability

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\,(Y_i = c_k \mid Y_{i-1} = c_l) \times \text{forward}(i-1, c_l) \right) \times P\,(X_i \mid Y_i = c_k)$$

$$\text{forward}(N+1, \mathit{STOP})$$

$$= P\,(Y_{N+1} = \mathit{STOP}, X_1 = x_1, \ldots, X_N = x_N, X_{N+1})$$

$$= P(X)$$

$$P(X) = \text{forward}(N+1, \mathit{STOP})$$

Decoding Problem

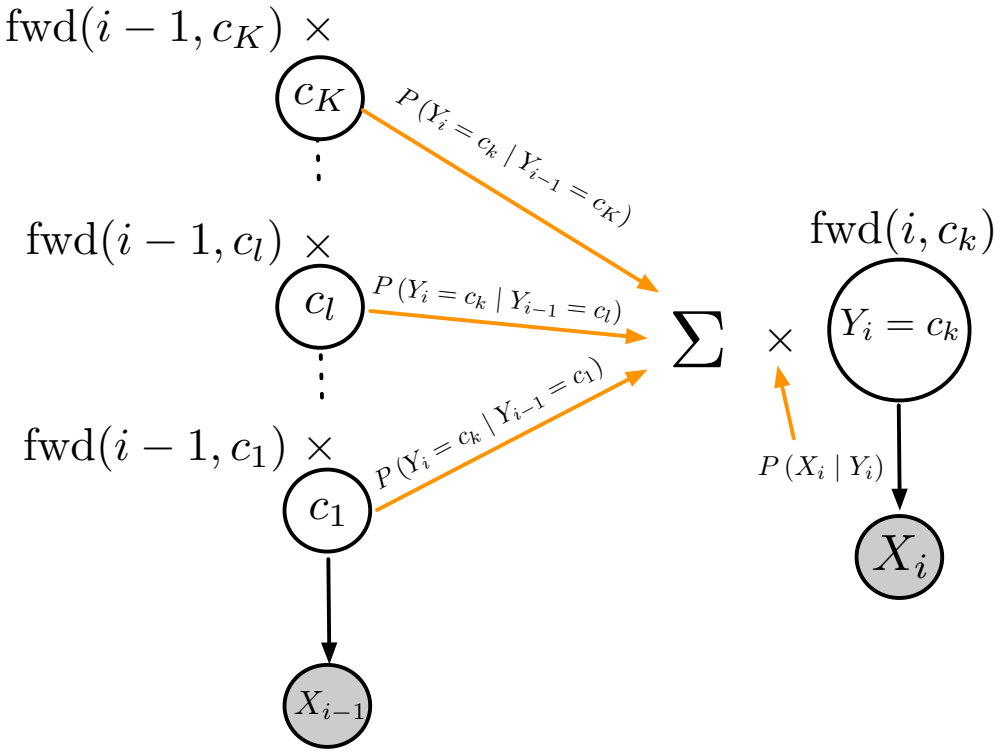
$$\text{Brute force approach } \sim O(K^N)$$

What is the complexity of forward probability calculations?

Decoding Problem

$$\text{Brute force approach } \sim O(K^N)$$

What is the complexity of forward probability calculations?

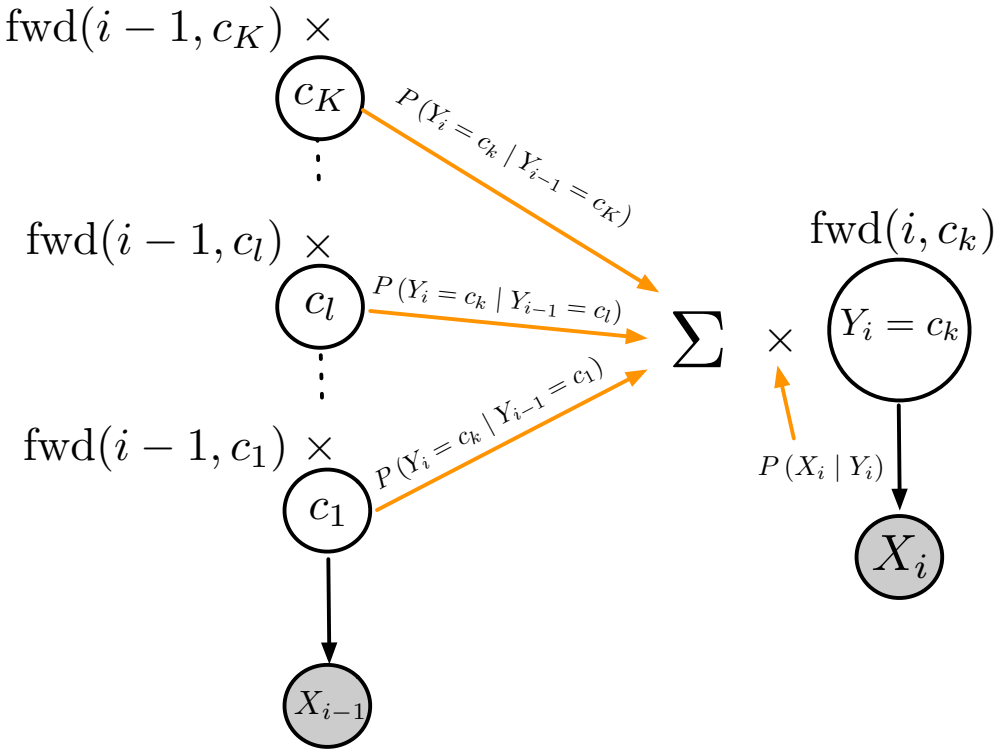


At any given time step, for every state we have information coming from K previous states. This accounts for O(K) operations

Decoding Problem

$$\text{Brute force approach } \sim O(K^N)$$

What is the complexity of forward probability calculations?



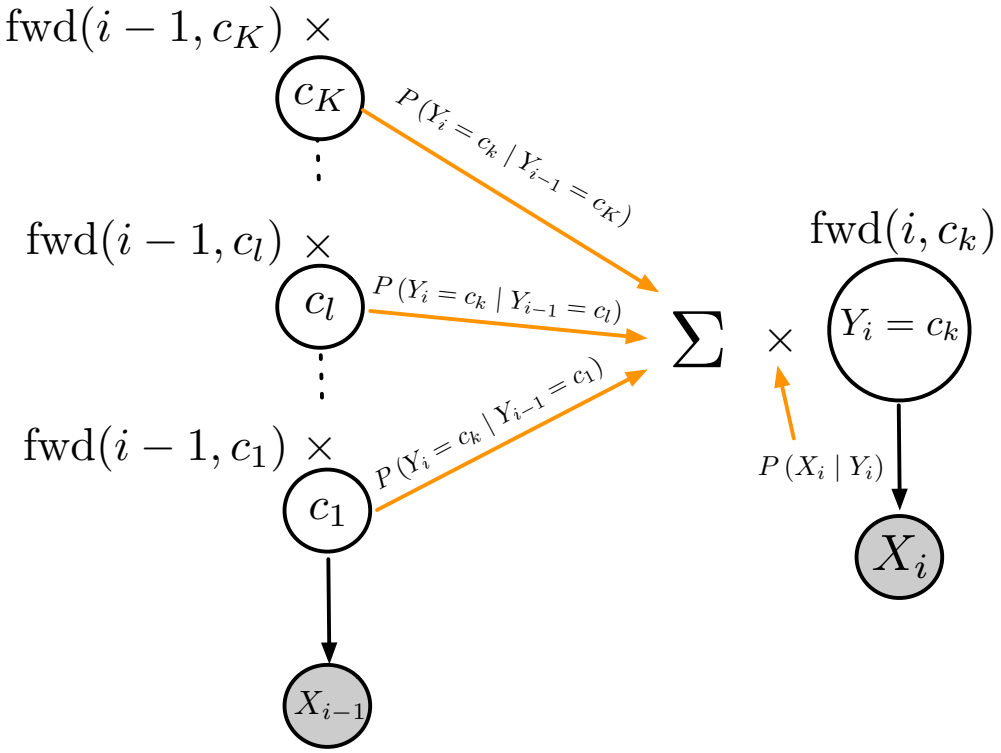
At any given time step, for every state we have information coming from K previous states. This accounts for O(K) operations

We do this for all states at a given time step. This makes it O(K²) operations.

Decoding Problem

$$\text{Brute force approach } \sim O(K^N)$$

What is the complexity of forward probability calculations?



At any given time step, for every state we have information coming from K previous states. This accounts for O(K) operations

We do this for all states at a given time step. This makes it $O(K^2)$ operations.

Repeating this over the whole sequence (length = N) leads to $O(N K^2)$ operations.

Decoding Problem

$$\text{Brute force approach } \sim O(K^N)$$

What is the complexity of forward probability calculations?

$$\text{Complexity of forward probability calculations} \sim O(NK^2)$$

Decoding Problem

$$\text{Brute force approach } \sim O(K^N)$$

$$\text{What is the complexity of forward probability calculations?}$$

$$\text{Complexity of forward probability calculations} \sim O(NK^2)$$

We have drastically reduced number of operations:
exponential → linear

Backward Probability

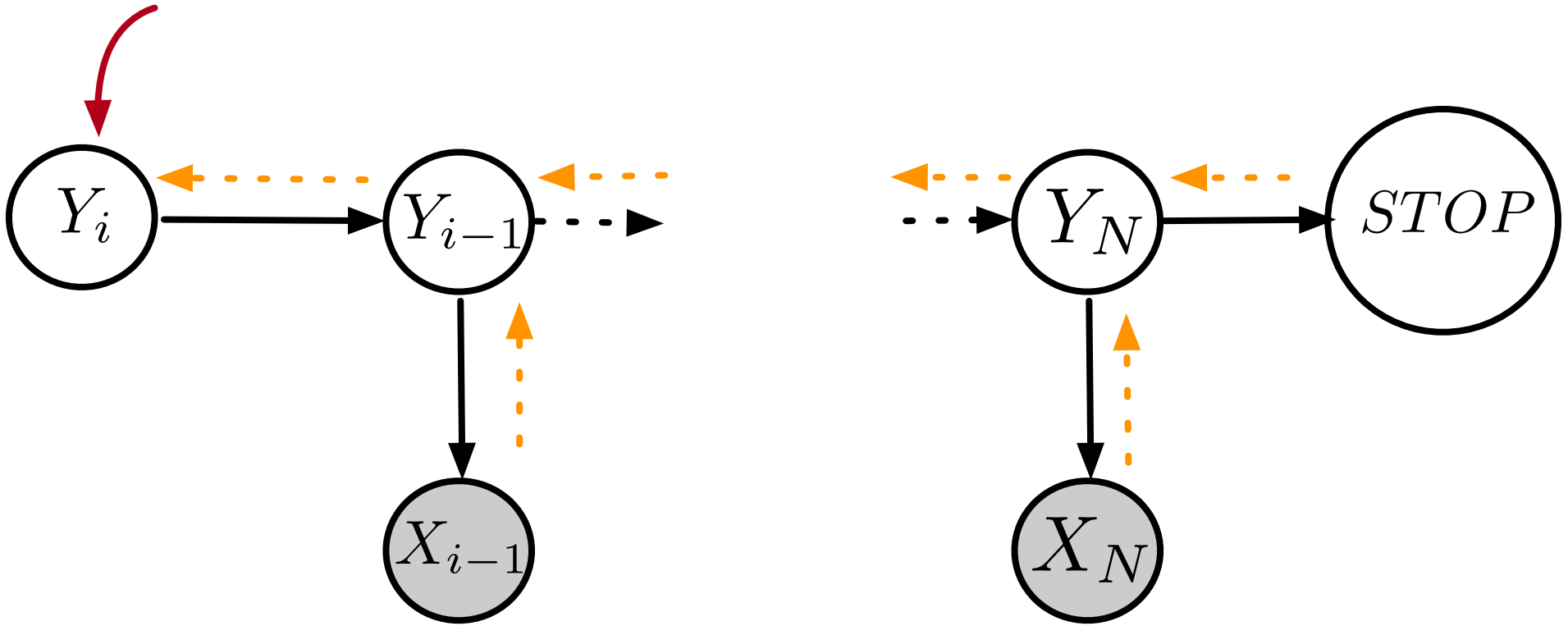
$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

Backward Probability

$$P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$



Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

$$\text{backward}(i, c_l) = \sum_{c_k} P\left(X_{i+1} = x_{i+1} \mid Y_{i+1} = c_k\right) \times \text{backward}(i+1, c_k) \times P\left(Y_{i+1} = c_k \mid Y_i = c_l\right)$$

Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

$$P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right) = P\left(X_{i+1:N} \mid Y_i\right)$$

Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

$$P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right) = P\left(X_{i+1:N} \mid Y_i\right)$$

$$P\left(X_{i+1:N} \mid Y_i\right)$$

Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

$$P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right) = P\left(X_{i+1:N} \mid Y_i\right)$$

$$P\left(X_{i+1:N} \mid Y_i\right)$$

$$= \sum_{Y_{i+1}} P\left(X_{i+1:N}, Y_{i+1} \mid Y_i\right)$$

Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

$$P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right) = P\left(X_{i+1:N} \mid Y_i\right)$$

$$P\left(X_{i+1:N} \mid Y_i\right)$$

$$= \sum_{Y_{i+1}} P\left(X_{i+1:N}, Y_{i+1} \mid Y_i\right)$$

$$= \sum_{Y_{i+1}} P\left(X_{i+1}, Y_{i+1}, X_{i+2:N} \mid Y_i\right)$$

Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

$$P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right) = P\left(X_{i+1:N} \mid Y_i\right)$$

$$P\left(X_{i+1:N} \mid Y_i\right)$$

$$= \sum_{Y_{i+1}} P\left(X_{i+1:N}, Y_{i+1} \mid Y_i\right)$$

$$= \sum_{Y_{i+1}} P\left(X_{i+1}, Y_{i+1}, X_{i+2:N} \mid Y_i\right)$$

$$= \sum_{Y_{i+1}} P\left(X_{i+1} \mid Y_{i+1}, X_{i+2:N}, Y_i\right) \times P\left(Y_{i+1}, X_{i+2:N}, Y_i\right)$$

Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

$$P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right) = P\left(X_{i+1:N} \mid Y_i\right)$$

$$P\left(X_{i+1:N} \mid Y_i\right)$$

$$= \sum_{Y_{i+1}} P\left(X_{i+1:N}, Y_{i+1} \mid Y_i\right)$$

$$= \sum_{Y_{i+1}} P\left(X_{i+1}, Y_{i+1}, X_{i+2:N} \mid Y_i\right)$$

$$= \sum_{Y_{i+1}} P\left(X_{i+1} \mid Y_{i+1}, X_{i+2:N}, Y_i\right) \times P\left(Y_{i+1}, X_{i+2:N} \mid Y_i\right)$$

$$= \sum_{Y_{i+1}} P\left(X_{i+1} \mid Y_{i+1}\right) \times P\left(X_{i+2:N} \mid Y_{i+1}\right) \times P\left(Y_{i+1} \mid Y_i\right)$$

Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

$$\text{backward}(i, c_l) = \sum_{c_k} P\left(X_{i+1} = x_{i+1} \mid Y_{i+1} = c_k\right) \times \text{backward}(i+1, c_k) \times P\left(Y_{i+1} = c_k \mid Y_i = c_l\right)$$

Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

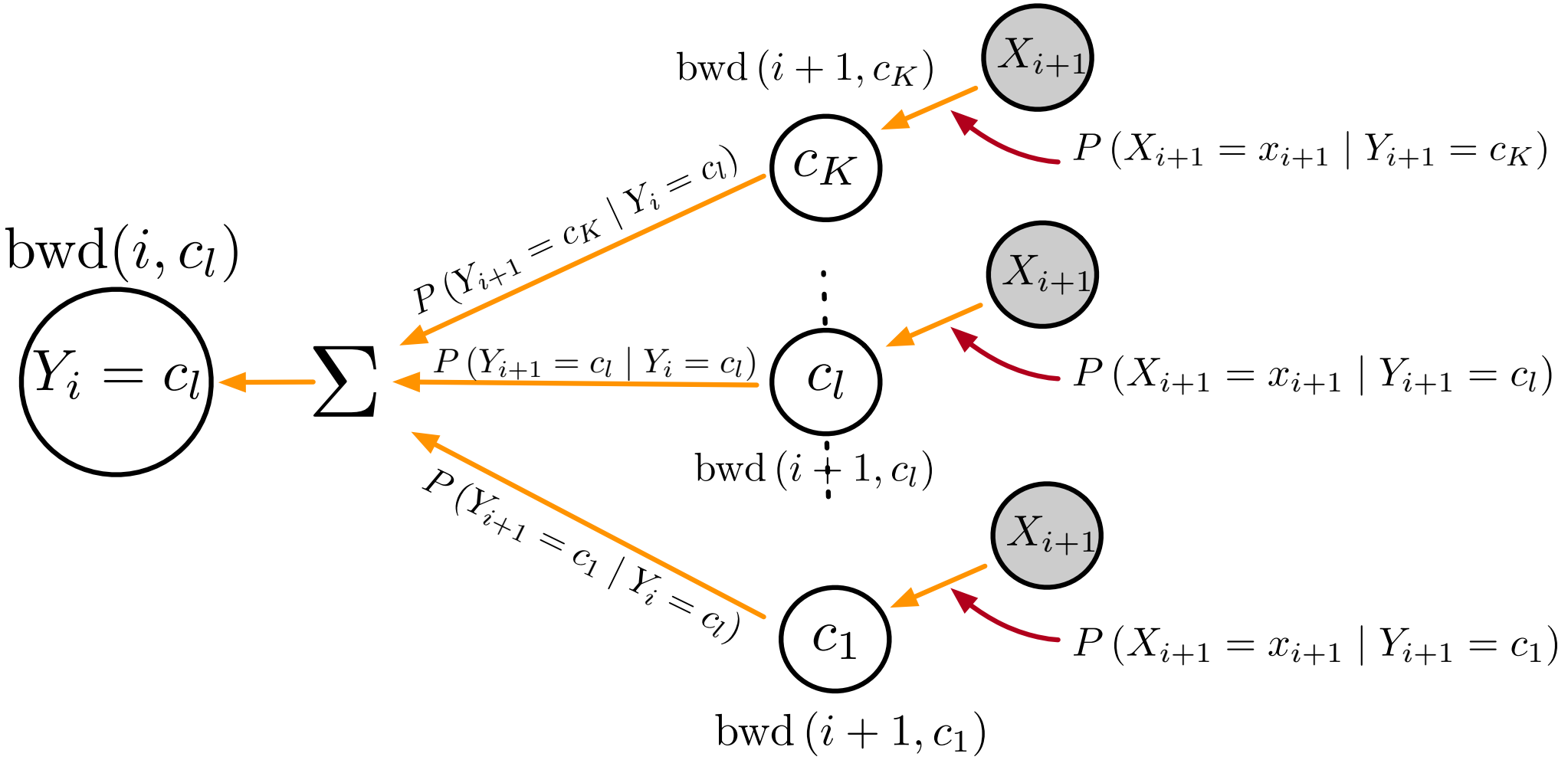
$$\text{backward}(i, c_l) = \sum_{c_k} P\left(X_{i+1} = x_{i+1} \mid Y_{i+1} = c_k\right) \times \text{backward}(i+1, c_k) \times P\left(Y_{i+1} = c_k \mid Y_i = c_l\right)$$

Emission Probability

Transition Probability

Backward Probability

$$\text{backward}(i, c_l) = \sum_{c_k} P\left(X_{i+1} = x_{i+1} \mid Y_{i+1} = c_k\right) \times \text{backward}(i+1, c_k) \times P\left(Y_{i+1} = c_k \mid Y_i = c_l\right)$$



Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

$$\text{backward}(i, c_l) = \sum_{c_k} P\left(X_{i+1} = x_{i+1} \mid Y_{i+1} = c_k\right) \times \text{backward}(i+1, c_k) \times P\left(Y_{i+1} = c_k \mid Y_i = c_l\right)$$

$$\text{backward}(0, START) = ?$$

Backward Probability

$$\text{backward}(i, c_l) := P\left(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l\right)$$

$$\text{backward}(i, c_l) = \sum_{c_k} P\left(X_{i+1} = x_{i+1} \mid Y_{i+1} = c_k\right) \times \text{backward}(i+1, c_k) \times P\left(Y_{i+1} = c_k \mid Y_i = c_l\right)$$

$$\text{backward}(0, START) = P(X)$$

Summary so far

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\,(Y_i = c_k \mid Y_{i-1} = c_l) \times \text{forward}(i-1, c_l) \right) \times P\,(X_i \mid Y_i = c_k)$$

Summary so far

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\,(Y_i = c_k \mid Y_{i-1} = c_l) \times \text{forward}(i-1, c_l) \right) \times P\,(X_i \mid Y_i = c_k)$$

$$\text{backward}(i, c_l) := P\,(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l)$$

$$\text{backward}(i, c_l) = \sum_{c_k} P\,(X_{i+1} = x_{i+1} \mid Y_{i+1} = c_k) \times \text{backward}(i+1, c_k) \times P\,(Y_{i+1} = c_k \mid Y_i = c_l)$$

Summary so far

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$\text{forward}\,(i, c_k) = \left(\sum_{c_l} P\,(Y_i = c_k \mid Y_{i-1} = c_l) \times \text{forward}(i-1, c_l) \right) \times P\,(X_i \mid Y_i = c_k)$$

$$\text{backward}(i, c_l) := P\,(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l)$$

$$\text{backward}(i, c_l) = \sum_{c_k} P\,(X_{i+1} = x_{i+1} \mid Y_{i+1} = c_k) \times \text{backward}(i+1, c_k) \times P\,(Y_{i+1} = c_k \mid Y_i = c_l)$$

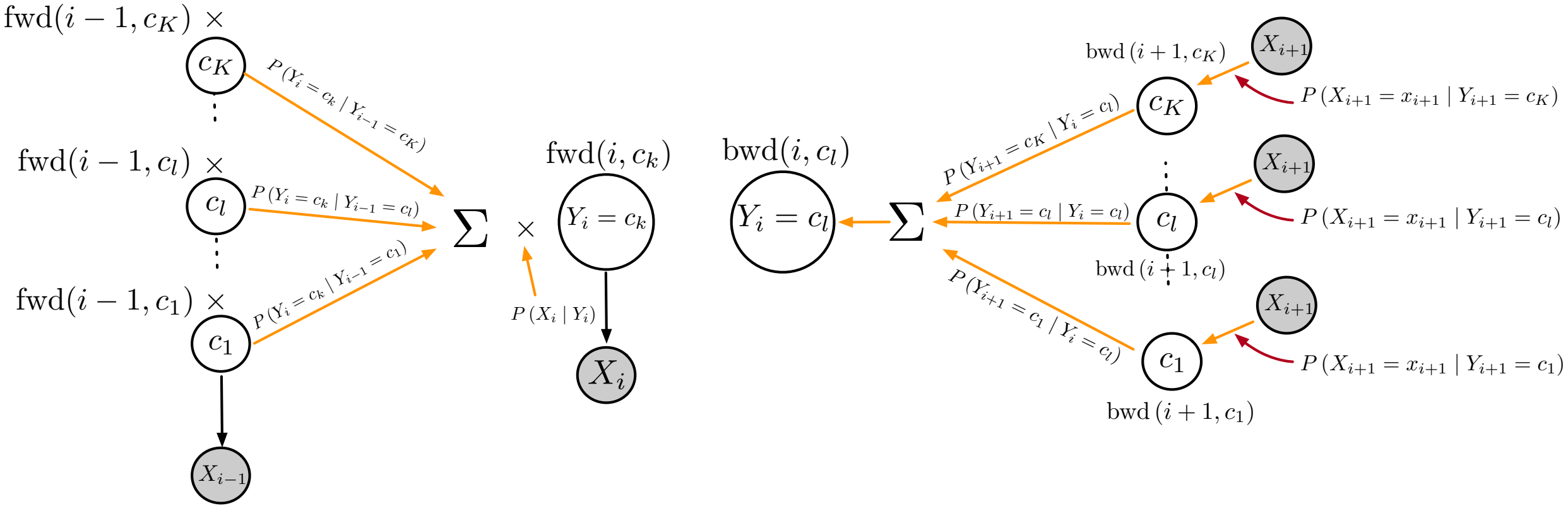
$$P(X) = \text{forward}(N+1, STOP)$$

$$\text{backward}(0, START) = P(X)$$

Summary so far

$$\text{forward}\,(i, c_k) := P\,(Y_i = c_k, X_1 = x_1, \ldots, X_i = x_i)$$

$$\text{backward}(i, c_l) := P\,(X_{i+1} = x_{i+1}, \ldots, X_N = x_N \mid Y_i = c_l)$$



Forward Backward

$$P(X) = P\left(X_1, \ldots, X_i, X_{i+1}, \ldots X_N\right)$$

Forward Backward

$$P(X) = P\left(X_1, \ldots, X_i, X_{i+1}, \ldots X_N\right)$$
$$= \sum_{c_l} P\left(X_1, \ldots, X_i, X_{i+1}, \ldots X_N, Y_i = c_l\right)$$

Forward Backward

$$
\begin{aligned}
P(X) &= P\left(X_1, \ldots, X_i, X_{i+1}, \ldots X_N\right) \\
&= \sum_{c_l} P\left(X_1, \ldots, X_i, X_{i+1}, \ldots X_N, Y_i = c_l\right) \\
&= \sum_{c_l} P\left(X_{i+1}, \ldots X_N \mid Y_i = c_l\right) \times P\left(Y_i = c_l, X_1, \ldots, X_i\right)
\end{aligned}
$$

Forward Backward

$$
\begin{aligned}
P(X) &= P\left(X_1, \ldots, X_i, X_{i+1}, \ldots X_N\right) \\
&= \sum_{c_l} P\left(X_1, \ldots, X_i, X_{i+1}, \ldots X_N, Y_i = c_l\right) \\
&= \sum_{c_l} P\left(X_{i+1}, \ldots X_N \mid Y_i = c_l\right) \times P\left(Y_i = c_l, X_1, \ldots, X_i\right) \\
&= \sum_{c_l} \text{backward}\left(i, c_l\right) \times \text{forward}\left(i, c_l\right)
\end{aligned}
$$

Forward Backward

$$P(X) = P(X_1, \ldots, X_i, X_{i+1}, \ldots X_N)$$

$$= \sum_{c_l} P(X_1, \ldots, X_i, X_{i+1}, \ldots X_N, Y_i = c_l)$$

$$= \sum_{c_l} P(X_{i+1}, \ldots X_N \mid Y_i = c_l) \times P(Y_i = c_l, X_1, \ldots, X_i)$$

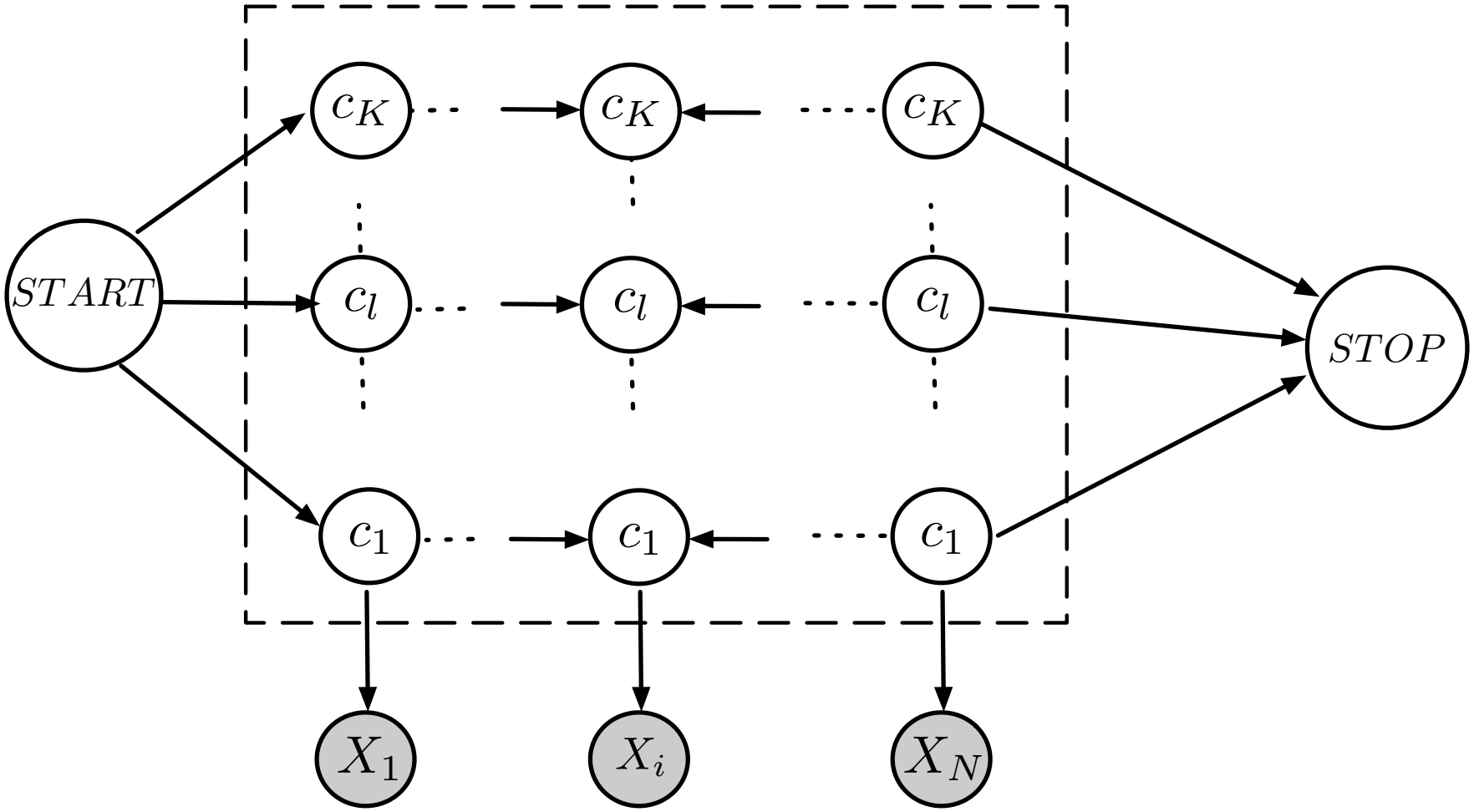
$$= \sum_{c_l} \text{backward}(i, c_l) \times \text{forward}(i, c_l)$$

Does not depend on *i*

$$P(X) = \sum_{c_l} \text{backward}(i, c_l) \times \text{forward}(i, c_l)$$

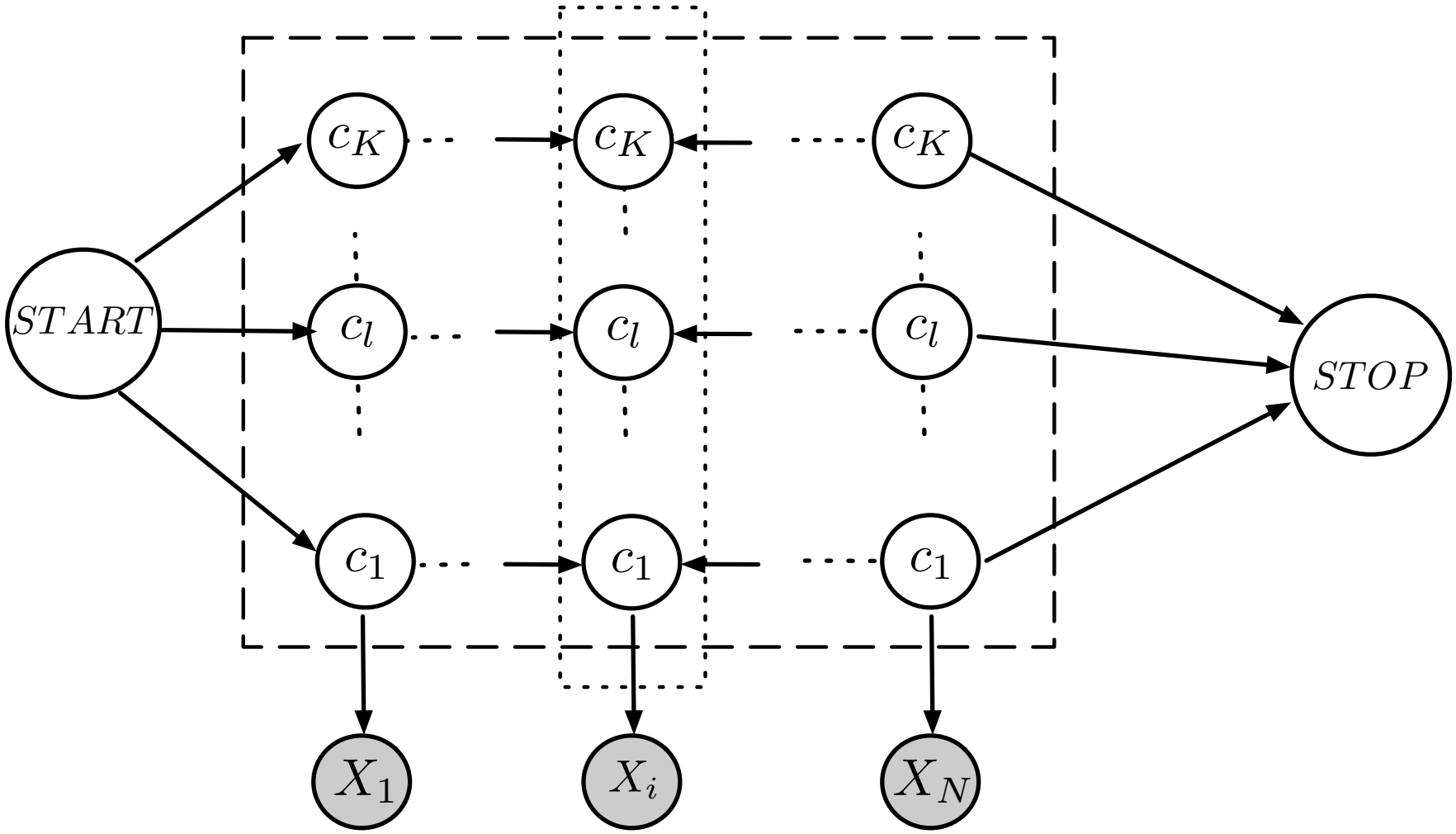
Forward Backward

$$P(X) = \sum_{c_l} \text{backward}\,(i, c_l) \times \text{forward}\,(i, c_l)$$



Forward Backward

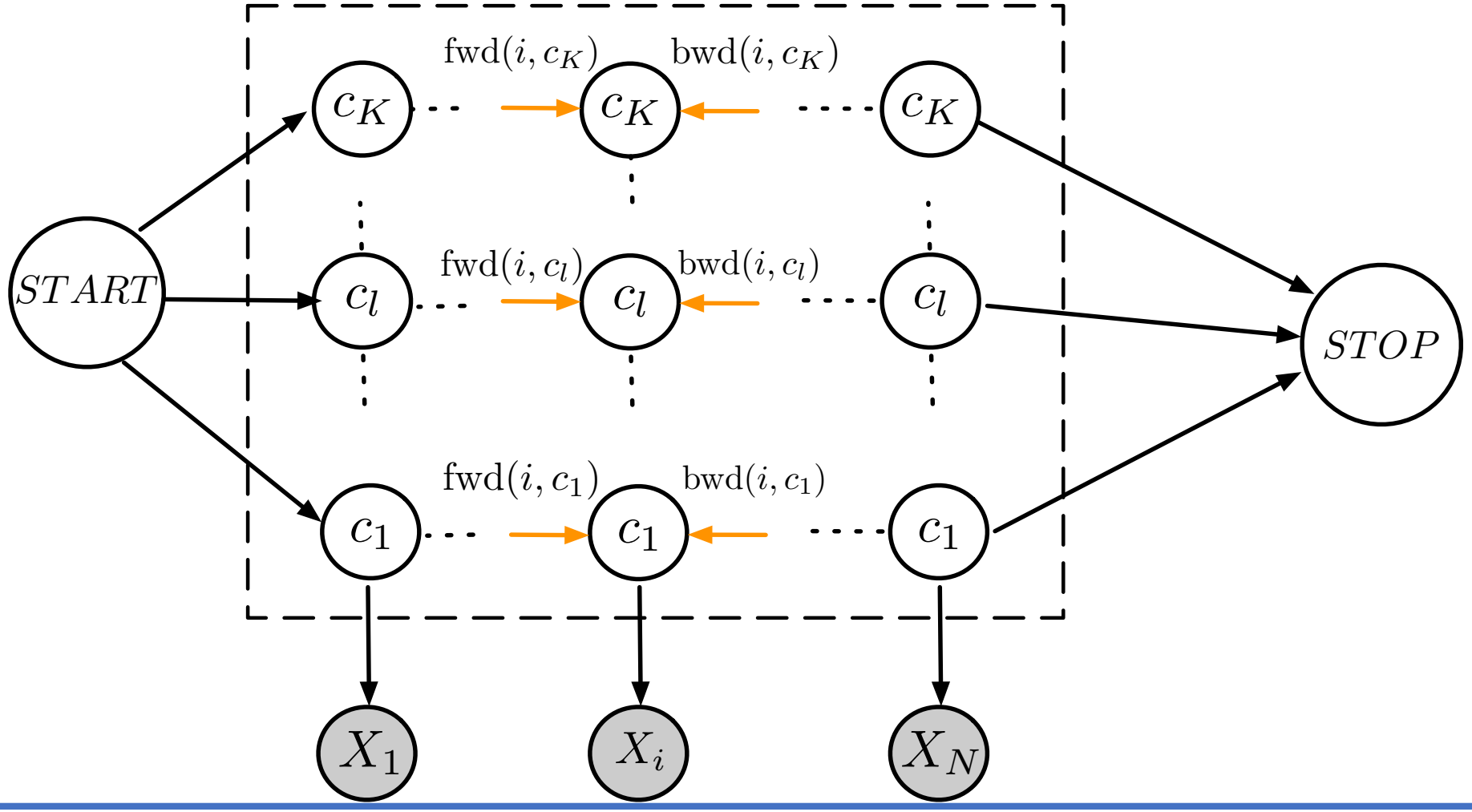
$$P(X) = \sum_{c_l} \text{backward}\,(i, c_l) \times \text{forward}\,(i, c_l)$$



Forward Backward

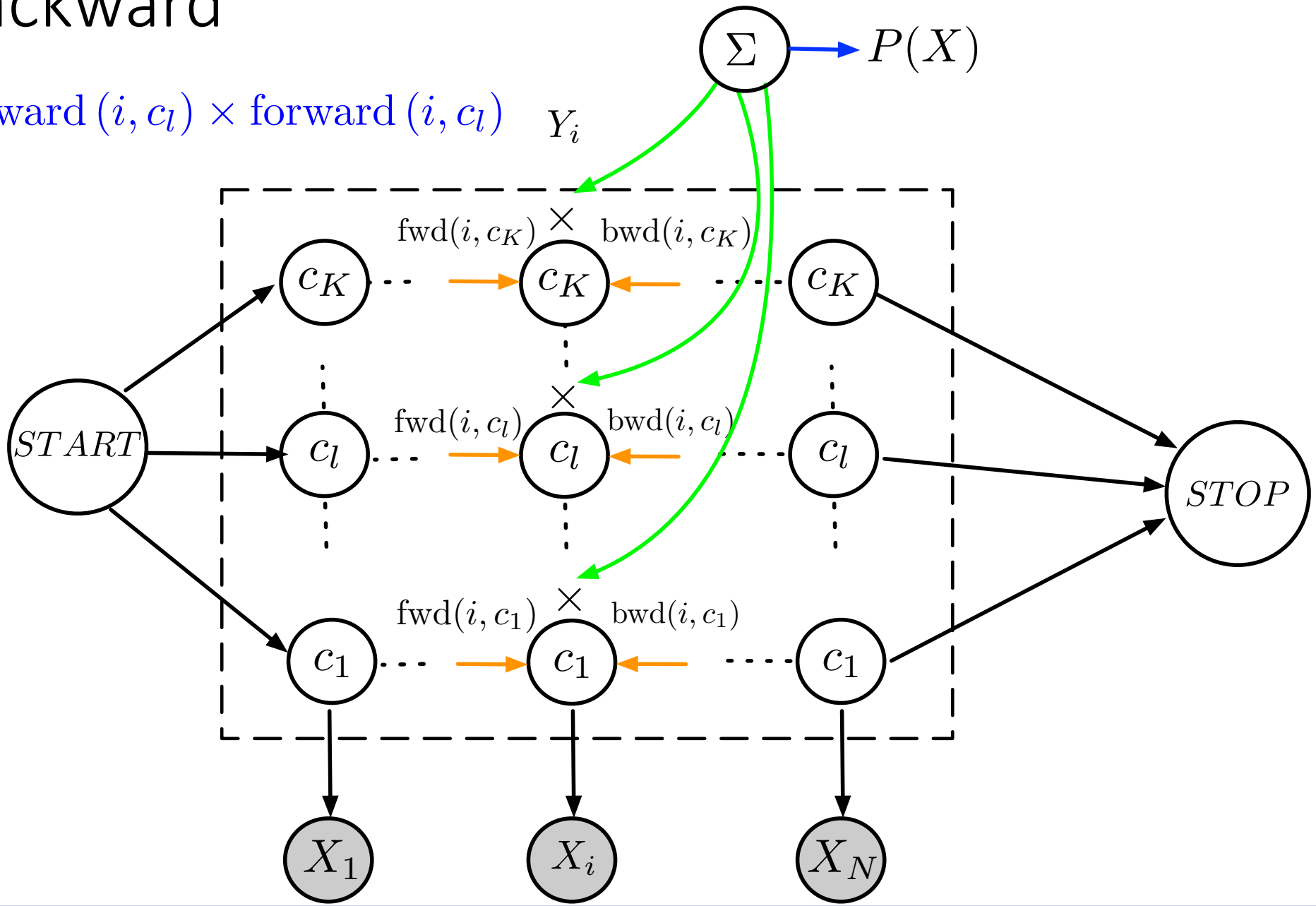
$$P(X) = \sum_{c_l} \text{backward}\,(i, c_l) \times \text{forward}\,(i, c_l)$$

Y_i



Forward Backward

$$P(X) = \sum_{c_l} \text{backward}\,(i, c_l) \times \text{forward}\,(i, c_l)$$



Computation

$$\text{Brute force approach } \sim O(K^N)$$

$$\text{What is the complexity of } P(X) \text{ calculation?}$$

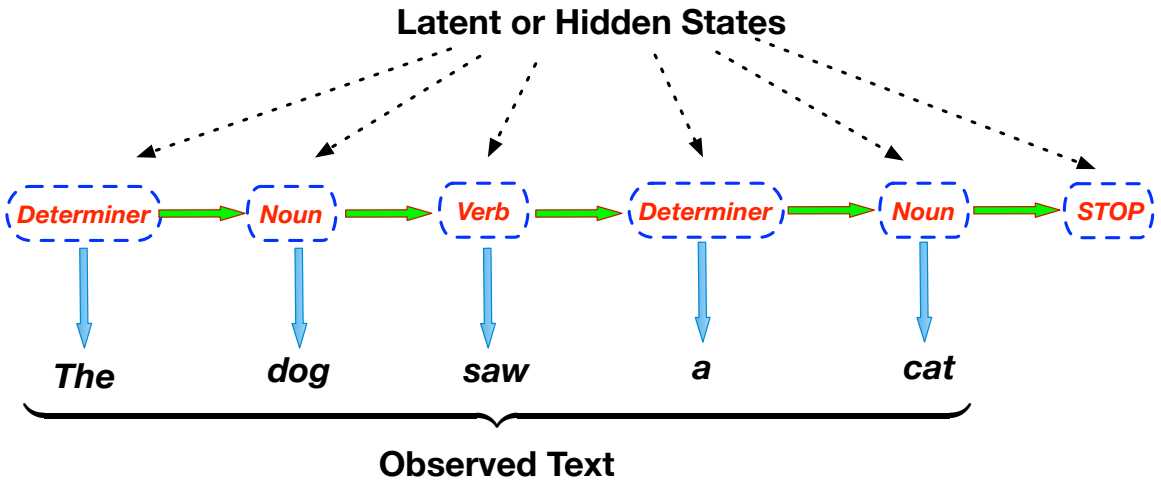
Computation

Brute force approach $\sim O(K^N)$

What is the complexity of $P(X)$ calculation?

Complexity of $P(X)$ calculation: $\sim O(NK^2)$

Hidden Markov Models (HMM)



Three Questions:

1. ***Learning Problem***: How do we estimate the parameters of distributions?
2. ***Decoding Problem***: Given the observed text what is the hidden POS sequence that best explains the observation?
3. ***Likelihood Problem***: Given the parameters of the distributions what is the probability of the observed sequence?

Decoding Problem

- Two ways possible
  - **Posterior Decoding**: Maximize probability of each state

$$y_i^* = \arg\max_{c_l} P\left(Y_i = c_l | X_1 = x_1, \ldots, X_N = x_N\right)$$

Decoding Problem

- Two ways possible
  - **Posterior Decoding**: Maximize probability of each state

$$y_i^* = \arg\max_{c_l} P\left(Y_i = c_l | X_1 = x_1, \ldots, X_N = x_N\right)$$

  - **Viterbi Decoding**: Maximize the probability of the entire sequence

$$y^* = \arg\max_{y=y_1 \ldots y_N} P\left(Y_1 = y_1, \ldots, Y_N = y_N | X_1 = x_1, \ldots, X_N = x_N\right)$$

$$= \arg\max_{y=y_1 \ldots y_N} P\left(Y_1 = y_1, \ldots, Y_N = y_N, X_1 = x_1, \ldots, X_N = x_N\right)$$

Posterior Decoding

$$y_i^* = \arg\max_{c_l} P\left(Y_i = c_l | X_1 = x_1, \ldots, X_N = x_N\right)$$

Posterior Decoding

$$y_i^* = \arg\max_{c_l} P\left(Y_i = c_l | X_1 = x_1, \ldots, X_N = x_N\right)$$

$$= \arg\max_{c_l} \frac{P\left(Y_i = c_l, X_1 = x_1, \ldots, X_N = x_N\right)}{P(X)}$$

Posterior Decoding

$$y_i^* = \arg\max_{c_l} P\left(Y_i = c_l | X_1 = x_1, \ldots, X_N = x_N\right)$$

$$= \arg\max_{c_l} \frac{P\left(Y_i = c_l, X_1 = x_1, \ldots, X_N = x_N\right)}{P(X)}$$

$$= \arg\max_{c_l} \frac{P\left(X_N = x_n, \ldots, X_{i+1} = x_{i+1} \mid Y_i = c_l\right) \times \left(Y_i = c_l, X_1 = x_1, \ldots, X_i = x_i\right)}{P(X)}$$

Posterior Decoding

$$y_i^* = \arg\max_{c_l} P\left(Y_i = c_l | X_1 = x_1, \ldots, X_N = x_N\right)$$

$$= \arg\max_{c_l} \frac{P\left(Y_i = c_l, X_1 = x_1, \ldots, X_N = x_N\right)}{P(X)}$$

$$= \arg\max_{c_l} \frac{P\left(X_N = x_n, \ldots, X_{i+1} = x_{i+1} \mid Y_i = c_l\right) \times \left(Y_i = c_l, X_1 = x_1, \ldots, X_i = x_i\right)}{P(X)}$$

$$= \arg\max_{c_l} \frac{\text{forward}(i, c_l) \times \text{backward}(i, c_l)}{P(X)}$$

Posterior Decoding

$$y_i^* = \arg\max_{c_l} P\left(Y_i = c_l | X_1 = x_1, \ldots, X_N = x_N\right)$$

$$= \arg\max_{c_l} \frac{P\left(Y_i = c_l, X_1 = x_1, \ldots, X_N = x_N\right)}{P(X)}$$

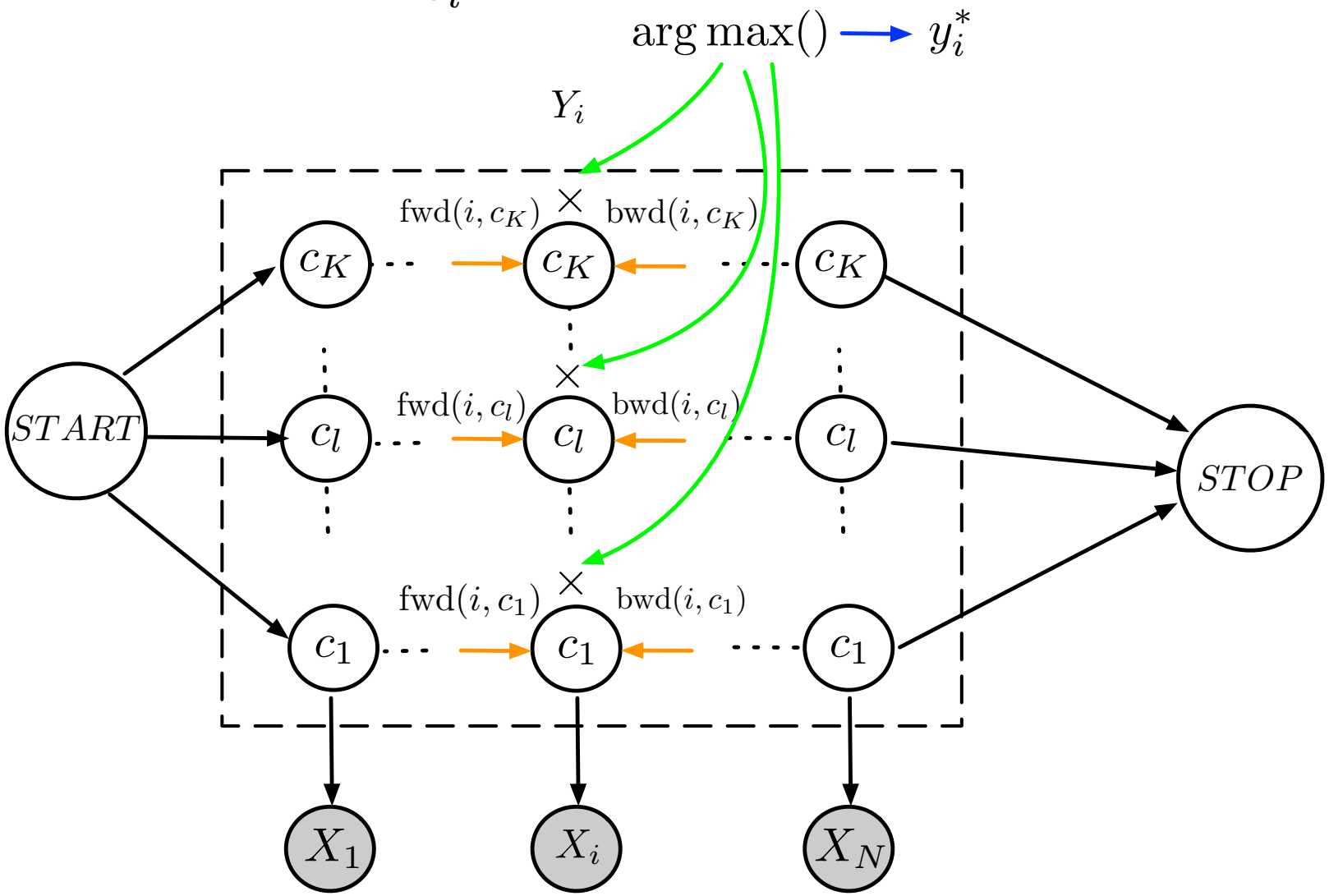
$$= \arg\max_{c_l} \frac{P\left(X_N = x_n, \ldots, X_{i+1} = x_{i+1} \mid Y_i = c_l\right) \times \left(Y_i = c_l, X_1 = x_1, \ldots, X_i = x_i\right)}{P(X)}$$

$$= \arg\max_{c_l} \frac{\text{forward}(i, c_l) \times \text{backward}(i, c_l)}{P(X)}$$

$$y_i^* = \arg\max_{c_l} \text{forward}(i, c_l) \times \text{backward}(i, c_l)$$

Posterior Decoding

$$y_i^* = \arg\max_{c_l} \text{forward}(i, c_l) \times \text{backward}(i, c_l)$$



References

1. Michael Collin's NLP Lecture Notes: http://www.cs.columbia.edu/~mcollins/hmms-spring2013.pdf

2. Chapter 6, Speech and Language Processing, Dan Jurafsky and James Martin

3. LxMLS Lab Guide: http://lxmls.it.pt/2016/LxMLS2016.pdf

QUIZ

- Suppose, you are given data about the activities that some person did over different days.
- We know that the activity that a person does on a given day is dependent on the weather on that day.
- We would like to predict the most likely weather pattern by observing the activity pattern.
- Just to simplify our problem, let us assume:
  - We have only 3 different activities: STUDY, SHOP, PLAY
  - We have only 2 possible weather conditions: SUNNY, RAIN

QUIZ

	STUDY (x_1)	SHOP (x_2)	PLAY (x_3)
SUNNY (c_1)	0.2	0.4	0.4
RAIN (c_2)	0.5	0.4	0.1

	SUNNY (c_1)	RAIN (c_2)
SUNNY (c_1)	0.6	0.4
RAIN (c_2)	0.5	0.5

	SUNNY (c_1)	RAIN (c_2)
START	0.8	0.2

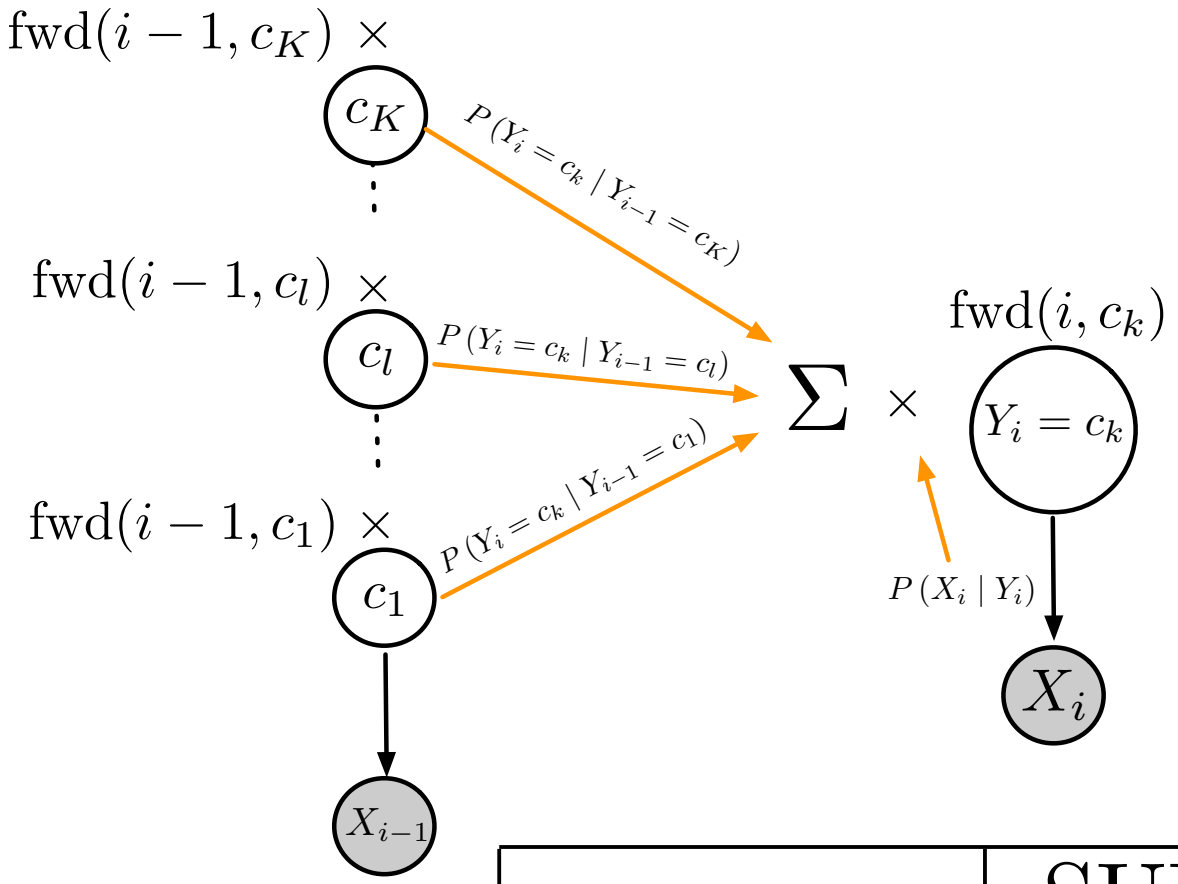
QUIZ

	STUDY (x_1)	SHOP (x_2)	PLAY (x_3)
SUNNY (c_1)	0.2	0.4	0.4
RAIN (c_2)	0.5	0.4	0.1

	SUNNY (c_1)	RAIN (c_2)
SUNNY (c_1)	0.6	0.4
RAIN (c_2)	0.5	0.5

	SUNNY (c_1)	RAIN (c_2)
START	0.8	0.2

Calculate the forward probabilities for the sequence PLAY, STUDY, PLAY



	STUDY (x_1)	SHOP (x_2)	PLAY (x_3)
SUNNY (c_1)	0.2	0.4	0.4
RAIN (c_2)	0.5	0.4	0.1

	SUNNY (c_1)	RAIN (c_2)
SUNNY (c_1)	0.6	0.4
RAIN (c_2)	0.5	0.5

	SUNNY (c_1)	RAIN (c_2)
START	0.8	0.2

Calculate the forward probabilities for the sequence: PLAY, STUDY, PLAY

	SUNNY (c_1)	RAIN (c_2)
$\text{fwd}(1, c_l)$	?	?
$\text{fwd}(2, c_l)$	?	?
$\text{fwd}(3, c_l)$	?	?