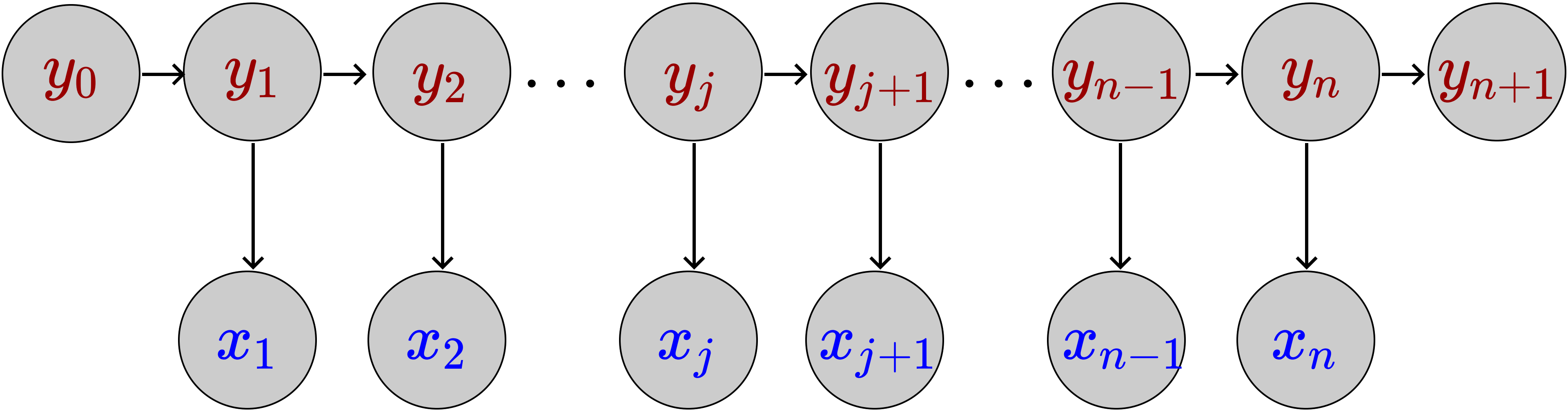
50.007
Machine Learning

Lu, Wei

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Hidden Markov Model (III)

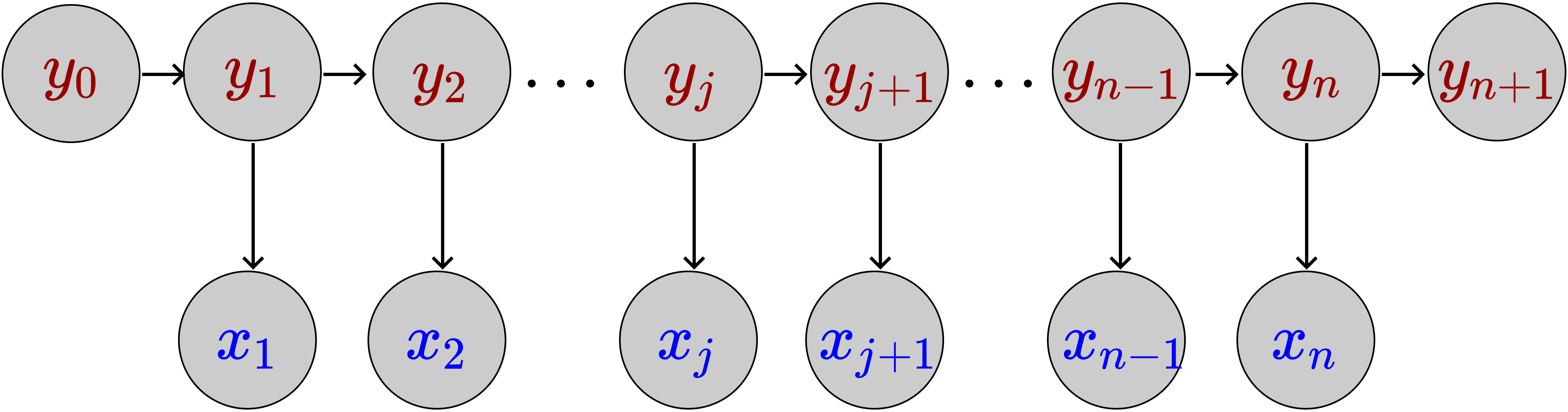
Hidden Markov Model
Parameterization



$$p(x_1, x_2, \ldots, x_{n-1}, x_n, y_0, y_1, y_2, \ldots, y_{n-1}, y_n, y_{n+1})$$

$$\underbrace{\prod_{j=0}^{n} a_{y_j, y_{j+1}}}_{\text{Transition probabilities}} \times \underbrace{\prod_{j=1}^{n} b_{y_j}(x_j)}_{\text{Emission probabilities}}$$

Hidden Markov Model
Supervised Learning



$$p(x_1, x_2, \ldots, x_{n-1}, x_n, y_0, y_1, y_2, \ldots, y_{n-1}, y_n, y_{n+1})$$

$$a_{u,v} = \frac{\mathrm{count}(u,v)}{\mathrm{count}(u)} \qquad b_u(o) = \frac{\mathrm{count}(u \to o)}{\mathrm{count}(u)}$$

Hidden Markov Model

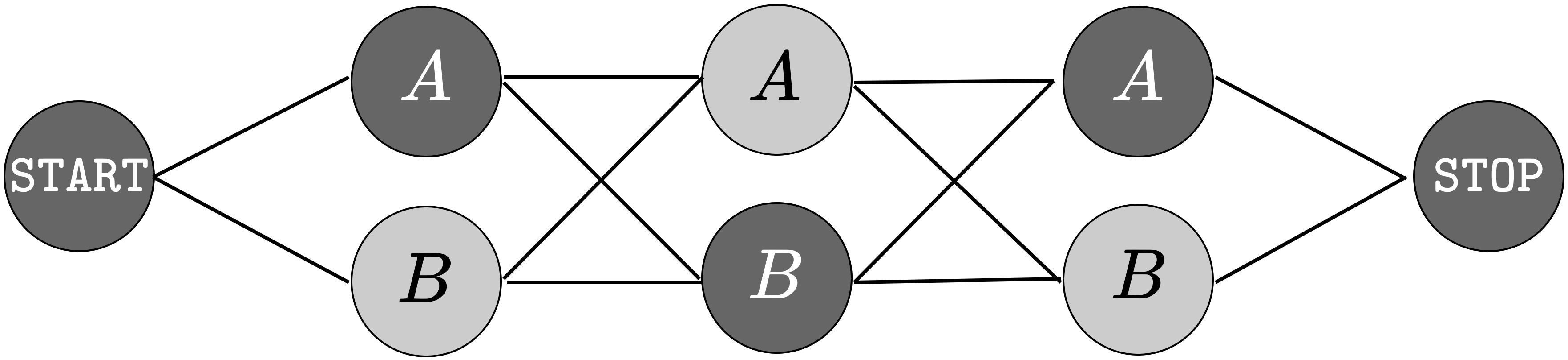
Decoding

$a_{u,v}$

$u \backslash v$	A	B	$\texttt{STOP}$
$\texttt{START}$	1.0	0.0	0.0
A	0.5	0.5	0.0
B	0.0	0.8	0.2

$b_u(o)$

$u \backslash o$	"the"	"dog"
A	0.9	0.1
B	0.1	0.9

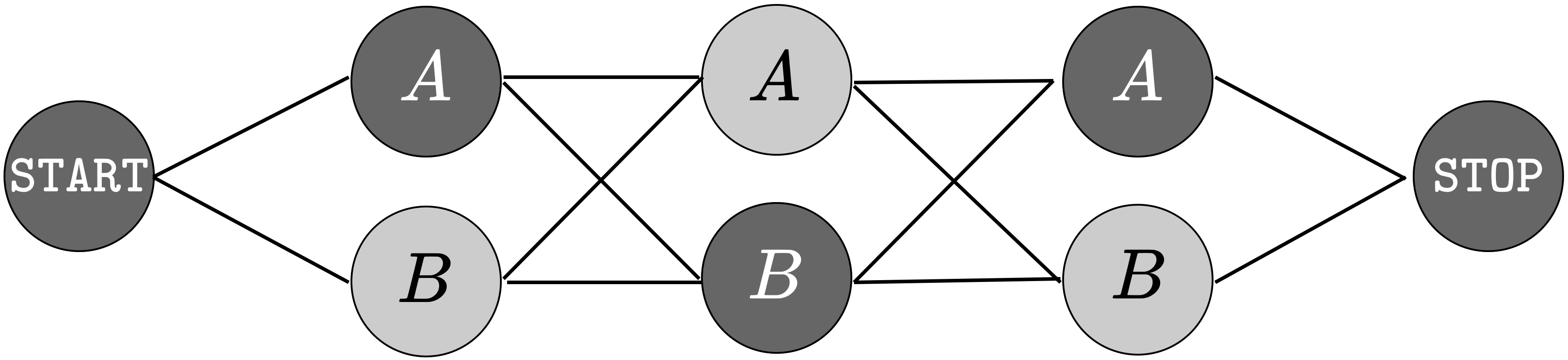


$$\mathbf{x} = \text{the}, \text{dog}, \text{the}$$

What is the most probable $\mathbf{y}$ sequence for the given $\mathbf{x}$ sequence?

Hidden Markov Model
Unsupervised Learning

We don't know the model parameters,
but only know there are two possible states: A, B.



$$\mathbf{x} = \text{the}, \text{dog}, \text{the}$$

What is the most probable $\mathbf{y}$ sequence for the given $\mathbf{x}$ sequence?

Question

How to solve the unsupervised learning problem for HMM?

Expectation Maximization

E-Step

Find for each input its membership

M-Step

Update the model parameters

Hard EM for HMM

E-Step

For each input sequence,

find its most probable output sequence.

This is the decoding procedure!

Hard EM for HMM

M-Step

Update the model parameters,

based on the training sequence pairs.

This is the supervised learning procedure!

Soft EM for HMM

E-Step

Run Viterbi, and then collect counts from each instance

M-Step

$$a_{u,v} = \frac{\text{count}(u,v)}{\text{count}(u)} \qquad b_u(o) = \frac{\text{count}(u \to o)}{\text{count}(u)}$$

Soft EM for HMM

E-Step

Run some algorithm to collect fractional counts from each instance

M-Step

$$a_{u,v} = \frac{\text{count}(u,v)}{\text{count}(u)} \qquad b_u(o) = \frac{\text{count}(u \to o)}{\text{count}(u)}$$

Soft EM for HMM

Finding the fractional count

A distribution over possible ys

$$\text{count}(u, v) = \sum_{i=1}^{m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}^{(i)})\text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \rightarrow v)$$

The number of times we see a transition from u to v in the sequence pair $(\mathbf{x}^{(\mathbf{i})}, \mathbf{y})$

Soft EM for HMM

$$\sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\mathrm{count}(\mathbf{x}, \mathbf{y}, u \to v)$$

Soft EM for HMM

$$\sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\mathrm{count}(\mathbf{x}, \mathbf{y}, u \to v)$$

$$= \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \sum_{j=0}^{n} \mathrm{count}(\mathbf{x}, \mathbf{y}, u \to v, j)$$

Soft EM for HMM

$$\sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\mathrm{count}(\mathbf{x}, \mathbf{y}, u \to v)$$

$$= \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \sum_{j=0}^{n} \mathrm{count}(\mathbf{x}, \mathbf{y}, u \to v, j)$$

$$= \sum_{j=0}^{n} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\mathrm{count}(\mathbf{x}, \mathbf{y}, u \to v, j)$$

Soft EM for HMM

$$\sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\text{count}(\mathbf{x}, \mathbf{y}, u \rightarrow v)$$

$$= \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \sum_{j=0}^{n} \text{count}(\mathbf{x}, \mathbf{y}, u \rightarrow v, j)$$

$$= \sum_{j=0}^{n} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\text{count}(\mathbf{x}, \mathbf{y}, u \rightarrow v, j)$$

This is an indicator function!

Soft EM for HMM

$$\sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\text{count}(\mathbf{x}, \mathbf{y}, u \rightarrow v, j)$$

Soft EM for HMM

$$\sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\mathrm{count}(\mathbf{x},\mathbf{y},u \rightarrow v, j)$$

$$= \sum_{y_0,\ldots,y_{n+1}} \Big[p(y_0,\ldots,y_j,y_{j+1},\ldots,y_{n+1}|\mathbf{x}) \\ \times \mathrm{count}(\mathbf{x},\mathbf{y},u \rightarrow v, j) \Big]$$

Soft EM for HMM

$$\sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\text{count}(\mathbf{x}, \mathbf{y}, u \to v, j)$$

$$= \sum_{y_0,\dots,y_{n+1}} \Big[p(y_0, \dots, y_j, y_{j+1}, \dots, y_{n+1}|\mathbf{x}) \\ \times \text{count}(\mathbf{x}, \mathbf{y}, u \to v, j) \Big]$$

$$= \sum_{y_0,\dots,y_{j-1},y_{j+2},\dots,y_{n+1}} \Big[p(y_0, \dots, y_{j-1}, \\ y_j = u, y_{j+1} = v, y_{j+2}, \dots, y_{n+1}|\mathbf{x}) \Big]$$

Soft EM for HMM

$$\sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\text{count}(\mathbf{x}, \mathbf{y}, u \to v, j)$$

$$= \sum_{y_0,\ldots,y_{n+1}} \Big[p(y_0, \ldots, y_j, y_{j+1}, \ldots, y_{n+1}|\mathbf{x}) \times \text{count}(\mathbf{x}, \mathbf{y}, u \to v, j)\Big]$$

$$= \sum_{y_0,\ldots,y_{j-1}, y_{j+2},\ldots,y_{n+1}} \Big[p(y_0, \ldots, y_{j-1}, y_j = u, y_{j+1} = v, y_{j+2}, \ldots, y_{n+1}|\mathbf{x})\Big]$$

$$= p(y_j = u, y_{j+1} = v|\mathbf{x})$$

Soft EM for HMM

$$\sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})\text{count}(\mathbf{x}, \mathbf{y}, u, j)$$

$$= \sum_{y_0,\ldots,y_{n+1}} \Big[p(y_0, \ldots, y_j, y_{j+1}, \ldots, y_{n+1}|\mathbf{x}) \times \text{count}(\mathbf{x}, \mathbf{y}, u, j)\Big]$$

$$= \sum_{y_0,\ldots,y_{j-1},y_{j+1},\ldots,y_{n+1}} \Big[p(y_0, \ldots, y_{j-1}, y_j = u, y_{j+1}, y_{j+2}, \ldots, y_{n+1}|\mathbf{x})\Big]$$

$$= p(y_j = u|\mathbf{x})$$

Inference in HMM

$$p(y_j = u|\mathbf{x}; \theta)$$

Model Parameters (α, β)

Inference in HMM

$$p(y_j = u|\mathbf{x}; \theta)$$

$$= \frac{p(y_j = u, \mathbf{x}; \theta)}{p(\mathbf{x}; \theta)}$$

Inference in HMM

$$p(y_j = u|\mathbf{x}; \theta)$$

$$= \frac{p(y_j = u, \mathbf{x}; \theta)}{p(\mathbf{x}; \theta)}$$

$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)}{p(x_1, x_2, \ldots, x_n; \theta)}$$

Inference in HMM

$$p(y_j = u | \mathbf{x}; \theta)$$

$$= \frac{p(y_j = u, \mathbf{x}; \theta)}{p(\mathbf{x}; \theta)}$$

$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)}{p(x_1, x_2, \ldots, x_n; \theta)}$$

What is the relation between the numerator and the denominator?

Inference in HMM

$$p(y_j = u | \mathbf{x}; \theta)$$

$$= \frac{p(y_j = u, \mathbf{x}; \theta)}{p(\mathbf{x}; \theta)}$$

$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)}{p(x_1, x_2, \ldots, x_n; \theta)}$$

$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)}{\sum_v p(x_1, x_2, \ldots, x_{k-1}, y_k = v, x_k, x_{k+1}, \ldots, x_n; \theta)}$$

Inference in HMM

$$p(y_j = u | \mathbf{x}; \theta)$$

$$= \frac{p(y_j = u, \mathbf{x}; \theta)}{p(\mathbf{x}; \theta)}$$

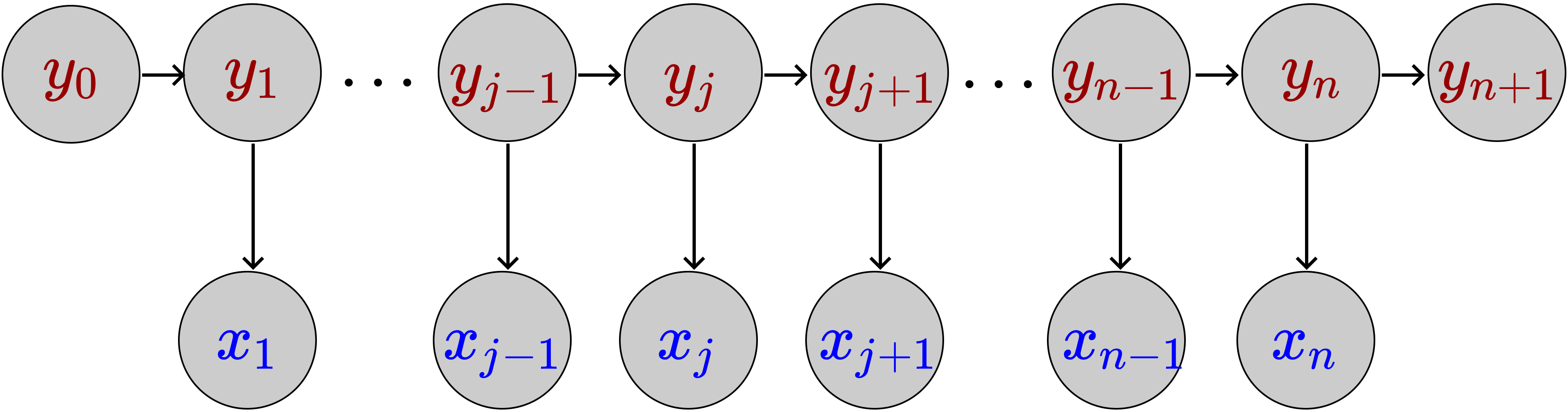
$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)}{p(x_1, x_2, \ldots, x_n; \theta)}$$

$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)}{\sum_v p(x_1, x_2, \ldots, x_{k-1}, y_k = v, x_k, x_{k+1}, \ldots, x_n; \theta)}$$

$$p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)$$

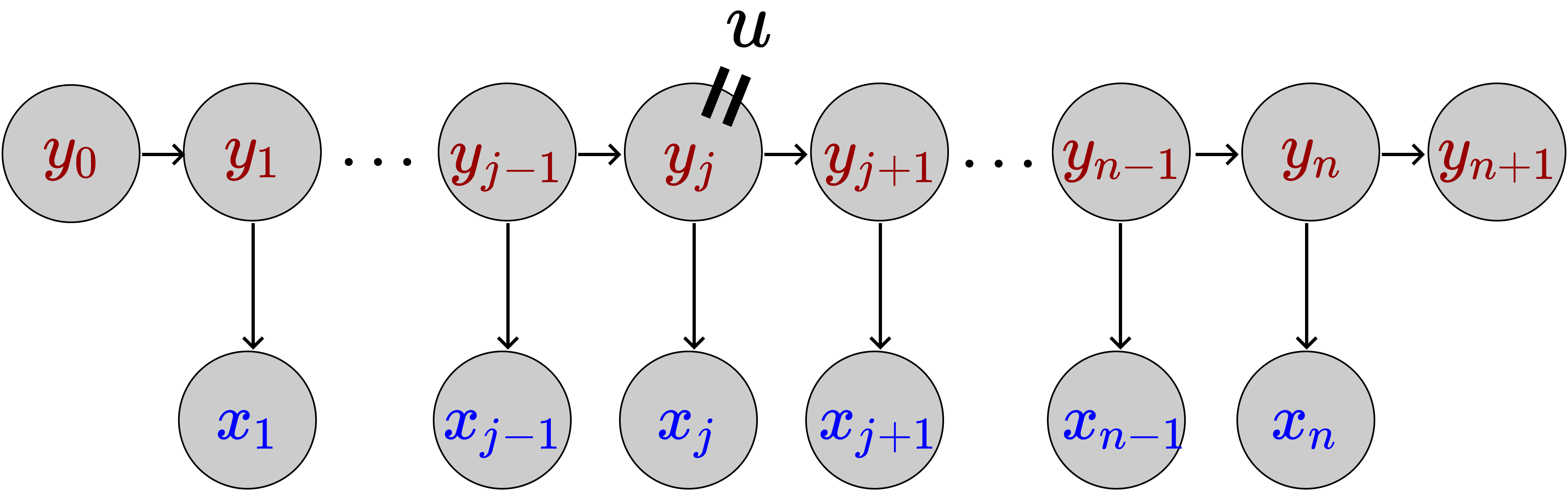
Now let us take a closer look at this joint probability.

Inference in HMM



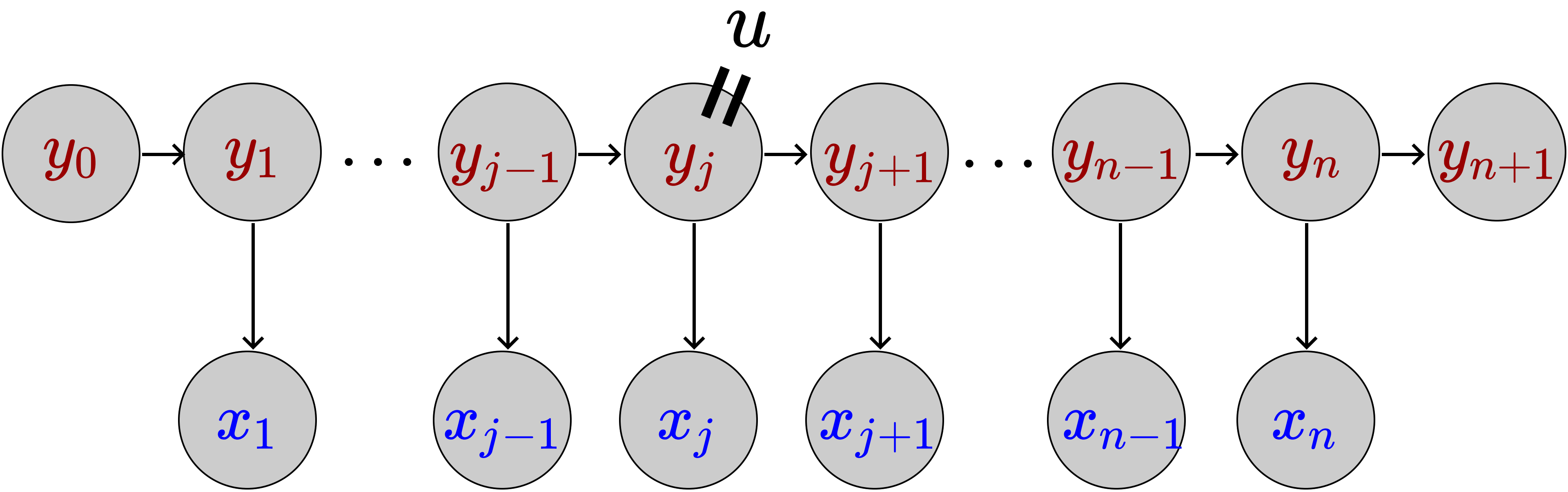
$$p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)$$

Inference in HMM



$$p(x_1, x_2, \dots, x_{j-1}, y_j = u, x_j, x_{j+1}, \dots, x_n; \theta)$$

Inference in HMM

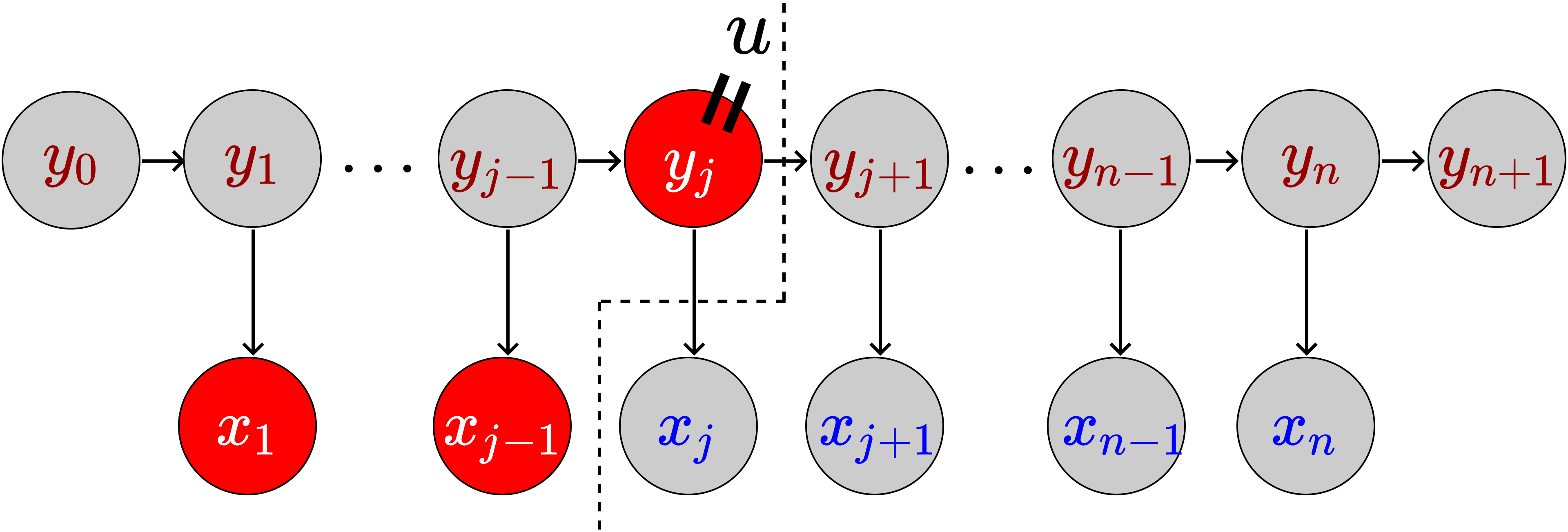


$$p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)$$

$$= \; p(x_1, x_2, \ldots, x_{j-1}, y_j = u; \theta)$$

$$p(x_j, x_{j+1}, \ldots, x_n | y_j = u; \theta)$$

Inference in HMM

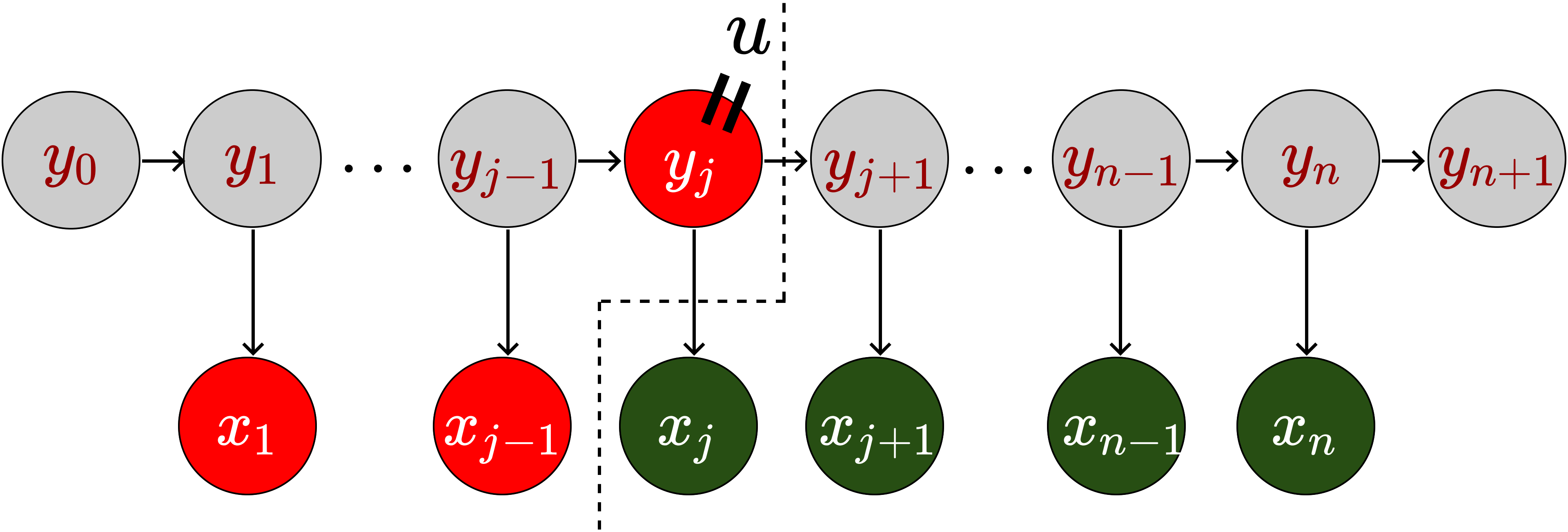


$$p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)$$

$$= \color{red}{p(x_1, x_2, \ldots, x_{j-1}, y_j = u; \theta)}$$

$$\times \; p(x_j, x_{j+1}, \ldots, x_n | y_j = u; \theta)$$

Inference in HMM

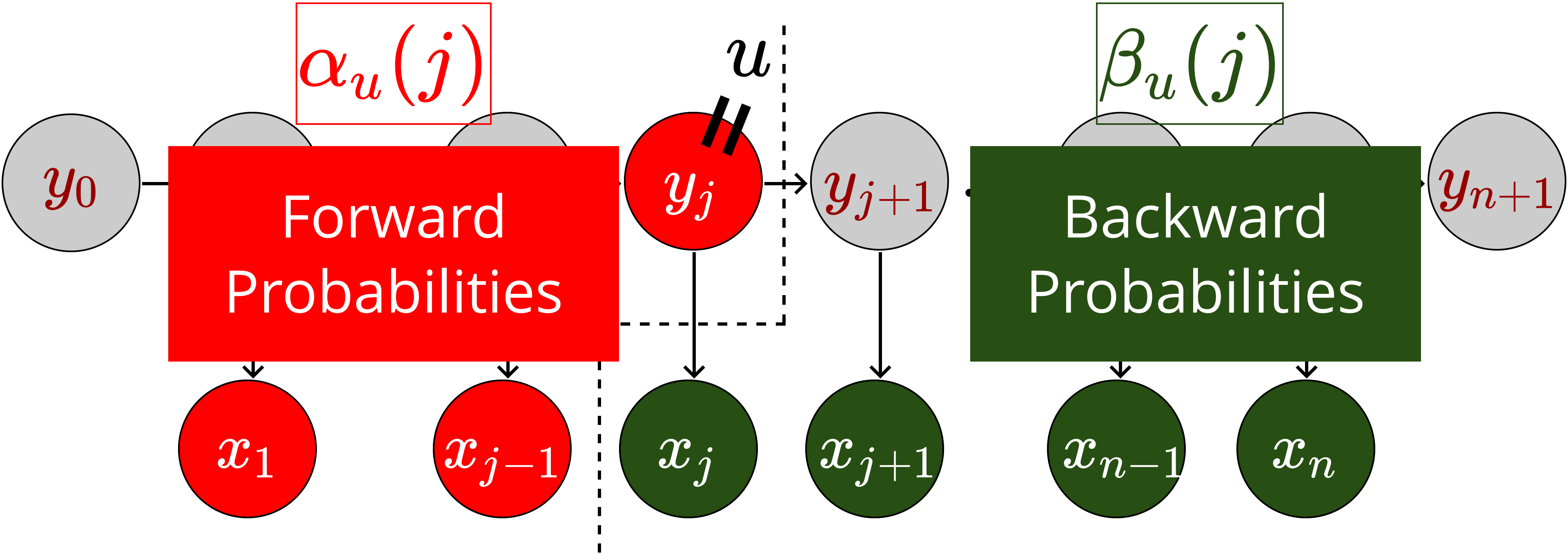


$$p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)$$

$$= p(x_1, x_2, \ldots, x_{j-1}, y_j = u; \theta)$$

$$\times\ p(x_j, x_{j+1}, \ldots, x_n | y_j = u; \theta)$$

Inference in HMM

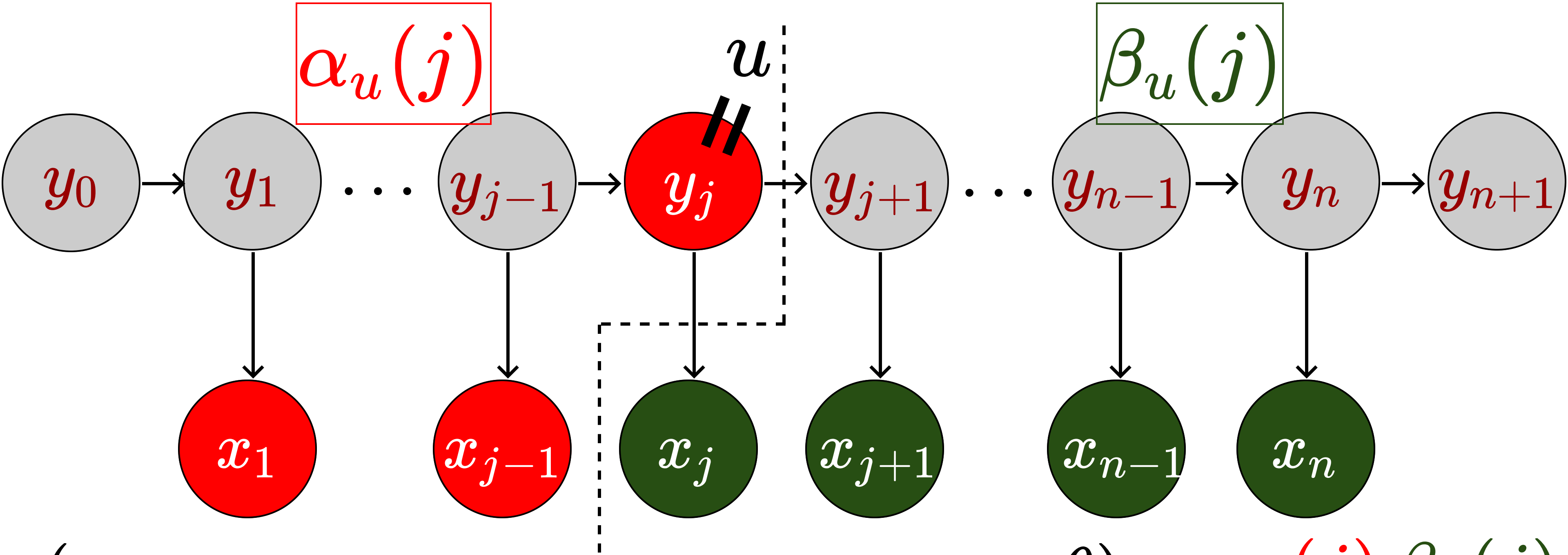


$$p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)$$

$$= p(x_1, x_2, \ldots, x_{j-1}, y_j = u; \theta) \quad \alpha_u(j)$$

$$\times\ p(x_j, x_{j+1}, \ldots, x_n | y_j = u; \theta) \quad \beta_u(j)$$

Inference in HMM

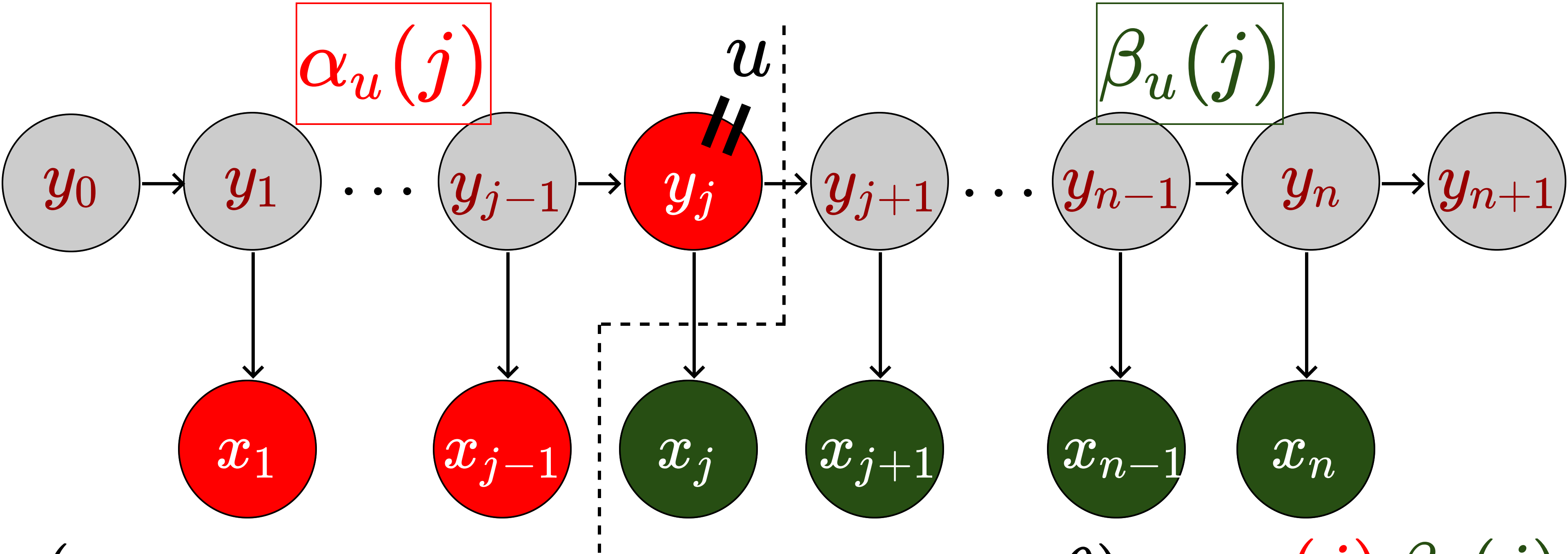


$$p(x_1, \ldots, x_{j-1}, y_j = u, x_j, \ldots, x_n; \theta) = \alpha_u(j)\, \beta_u(j)$$

$$p(y_j = u | \mathbf{x}; \theta)$$

$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)}{\sum_v p(x_1, x_2, \ldots, x_{k-1}, y_k = v, x_k, x_{k+1}, \ldots, x_n; \theta)}$$

Inference in HMM

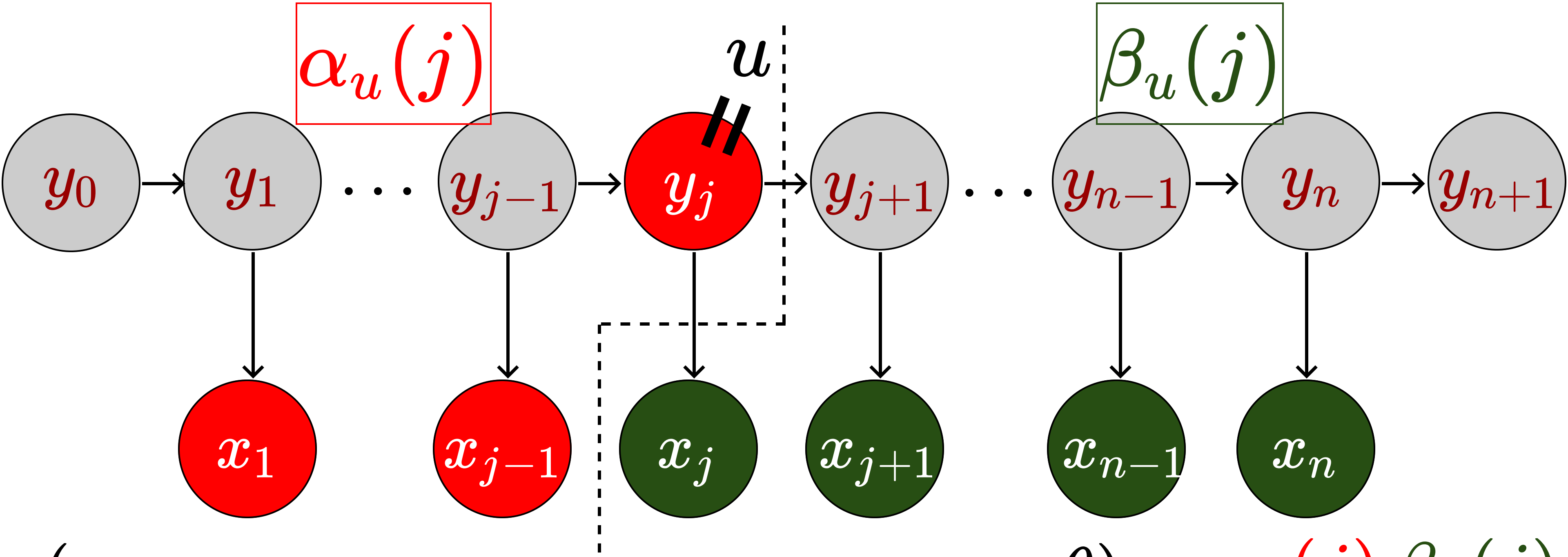


$$p(x_1, \ldots, x_{j-1}, y_j = u, x_j, \ldots, x_n; \theta) = \alpha_u(j)\, \beta_u(j)$$

$$p(y_j = u | \mathbf{x}; \theta)$$

$$= \frac{\alpha_u(j)\beta_u(j)}{\sum_v \alpha_v(k)\beta_v(k)}$$

Inference in HMM

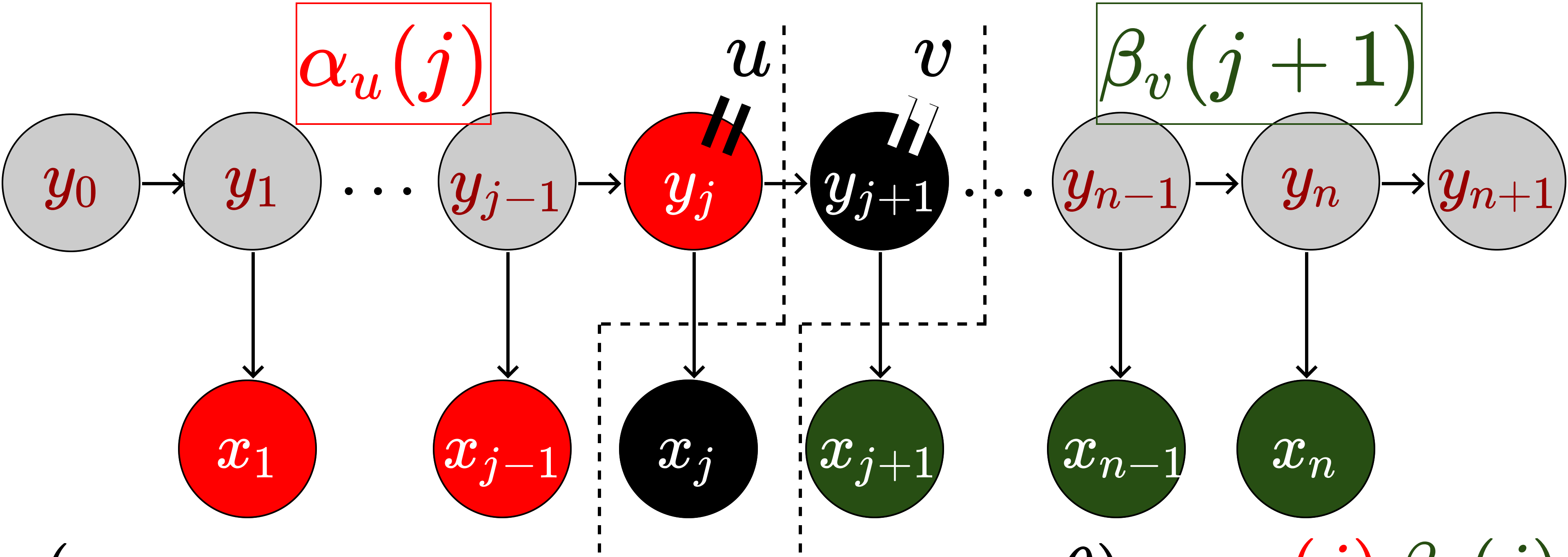


$$p(x_1, \ldots, x_{j-1}, y_j = u, x_j, \ldots, x_n; \theta) = \alpha_u(j)\, \beta_u(j)$$

$$p(y_j = u, y_{j+1} = v | \mathbf{x}; \theta)$$

$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, y_{j+1} = v, x_{j+1}, \ldots, x_n; \theta)}{\sum_v p(x_1, x_2, \ldots, x_{k-1}, y_k = v, x_k, x_{k+1}, \ldots, x_n; \theta)}$$

Inference in HMM



$$p(x_1, \ldots, x_{j-1}, y_j = u, x_j, \ldots, x_n; \theta) = \alpha_u(j)\, \beta_u(j)$$

$$p(y_j = u, y_{j+1} = v | \mathbf{x}; \theta)$$

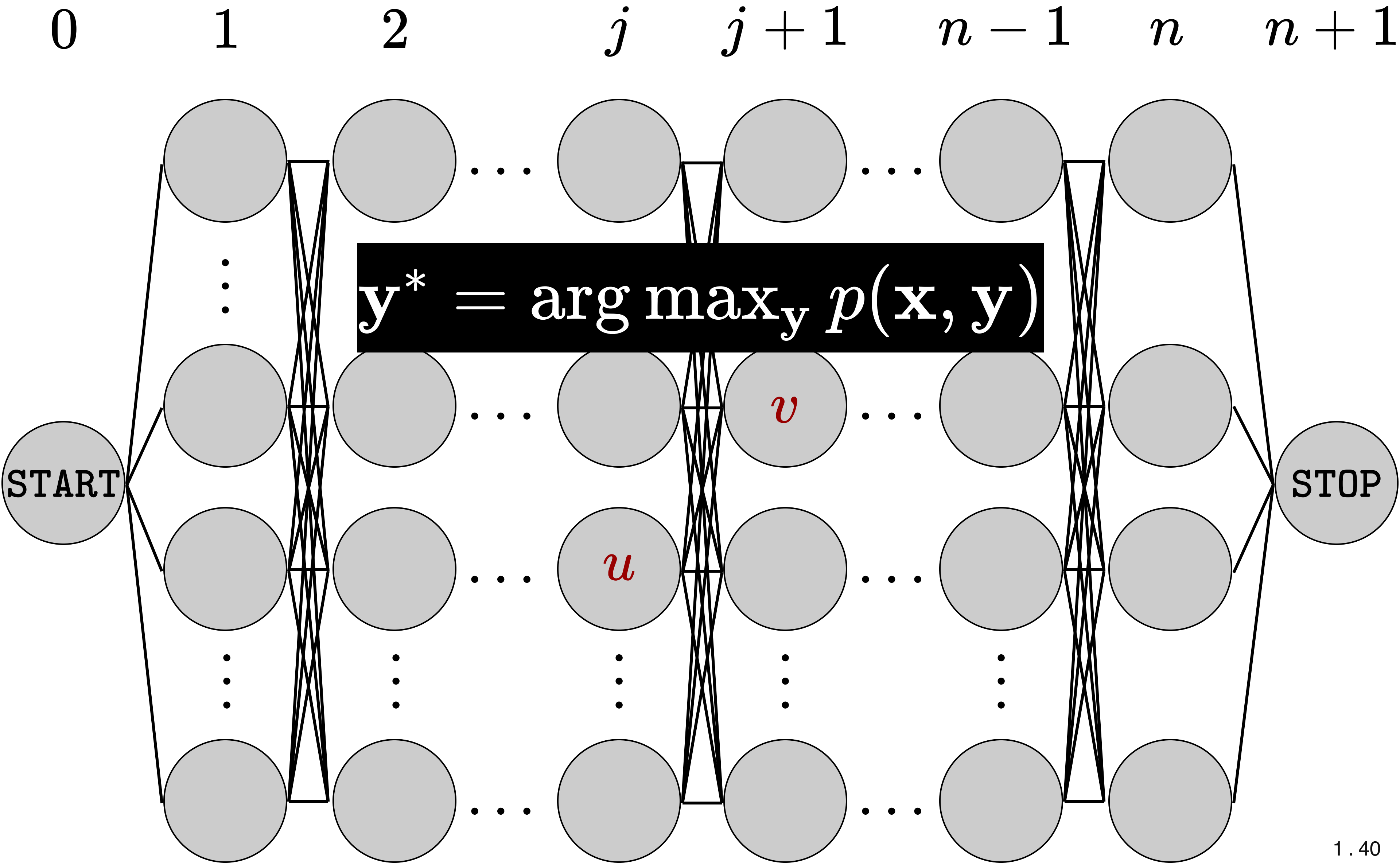
$$= \frac{\alpha_u(j) \cdot b_u(x_j) \cdot a_{u,v} \cdot \beta_v(j+1)}{\sum_v \alpha_v(k) \beta_v(k)}$$

Question

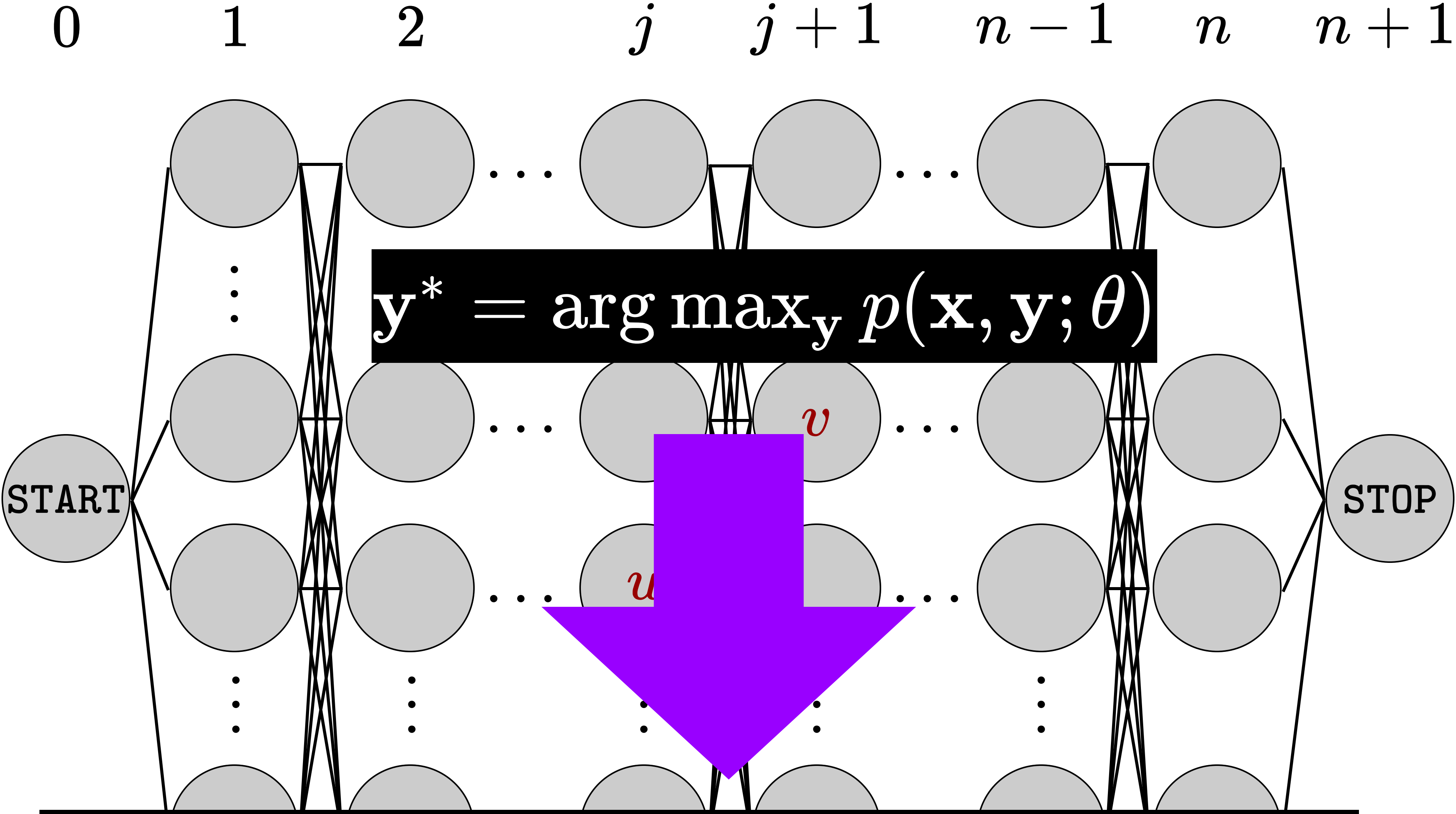
How to find an efficient procedure to calculate forward and backward probabilities?

Calculate forward/backward scores efficiently	→	Perform inference efficiently	→	Calculate the expected counts efficiently	→	Perform the soft-EM efficiently

Viterbi Algorithm



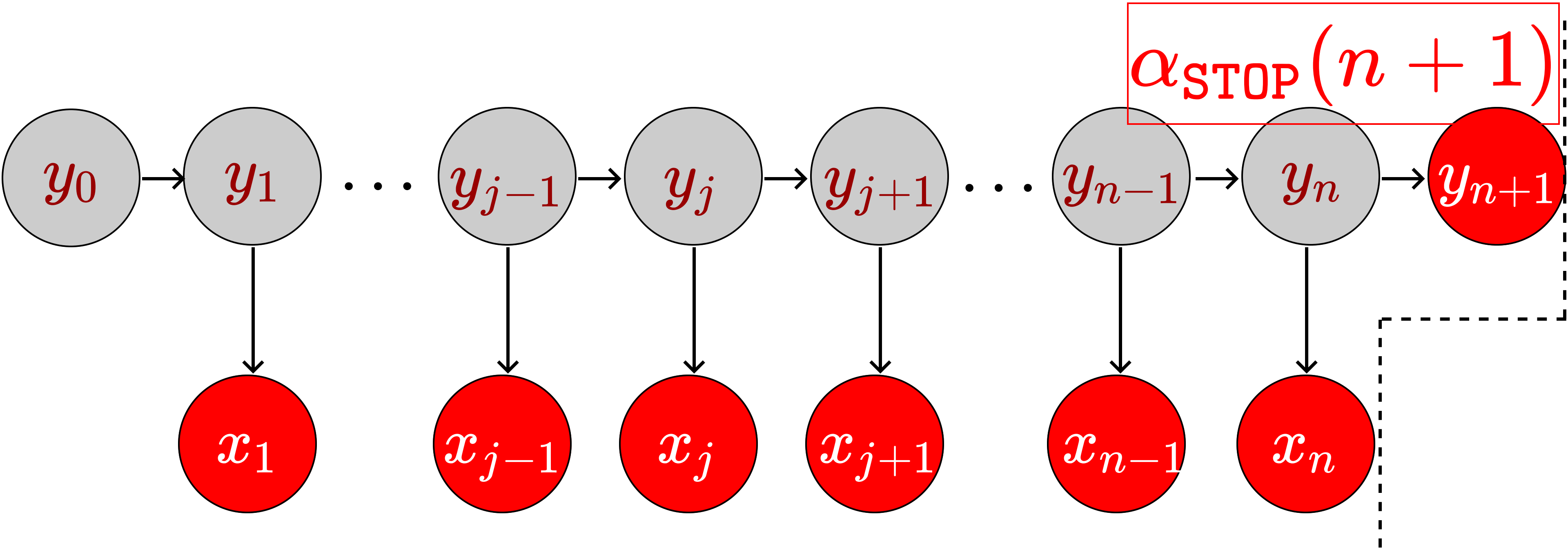
Viterbi Algorithm



$$\mathbf{y}^* = \arg\max_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}; \theta)$$

$$\max_{y_1,\ldots,y_n} p(x_1, \ldots, x_n, y_0, y_1, \ldots, y_n, y_{n+1}; \theta)$$

Inference in HMM

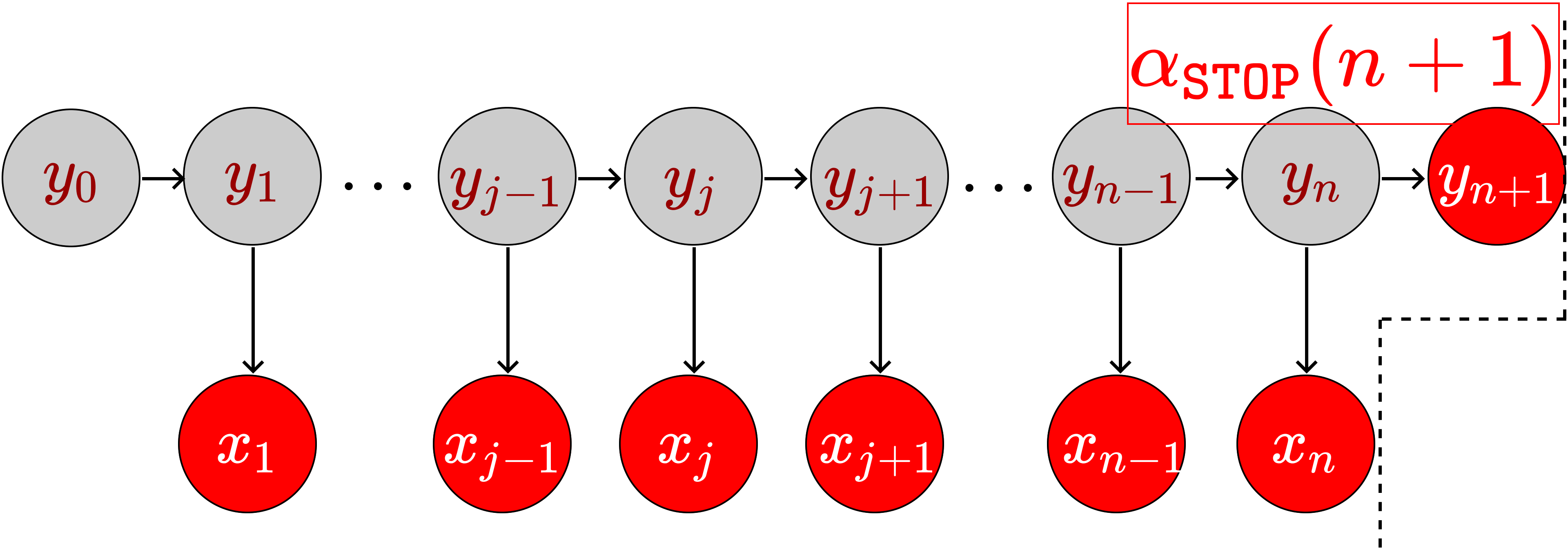


$\alpha_{\texttt{STOP}}(n+1)$

$$= \; p(x_1, \ldots, x_{j-1}, x_j, \ldots, x_n; \theta)$$

$$= \; \sum_{y_1,\ldots,y_n} p(x_1, \ldots, x_n, y_0, y_1, \ldots, y_n, y_{n+1}; \theta)$$

Inference in HMM



$\alpha_{\texttt{STOP}}(n+1)$

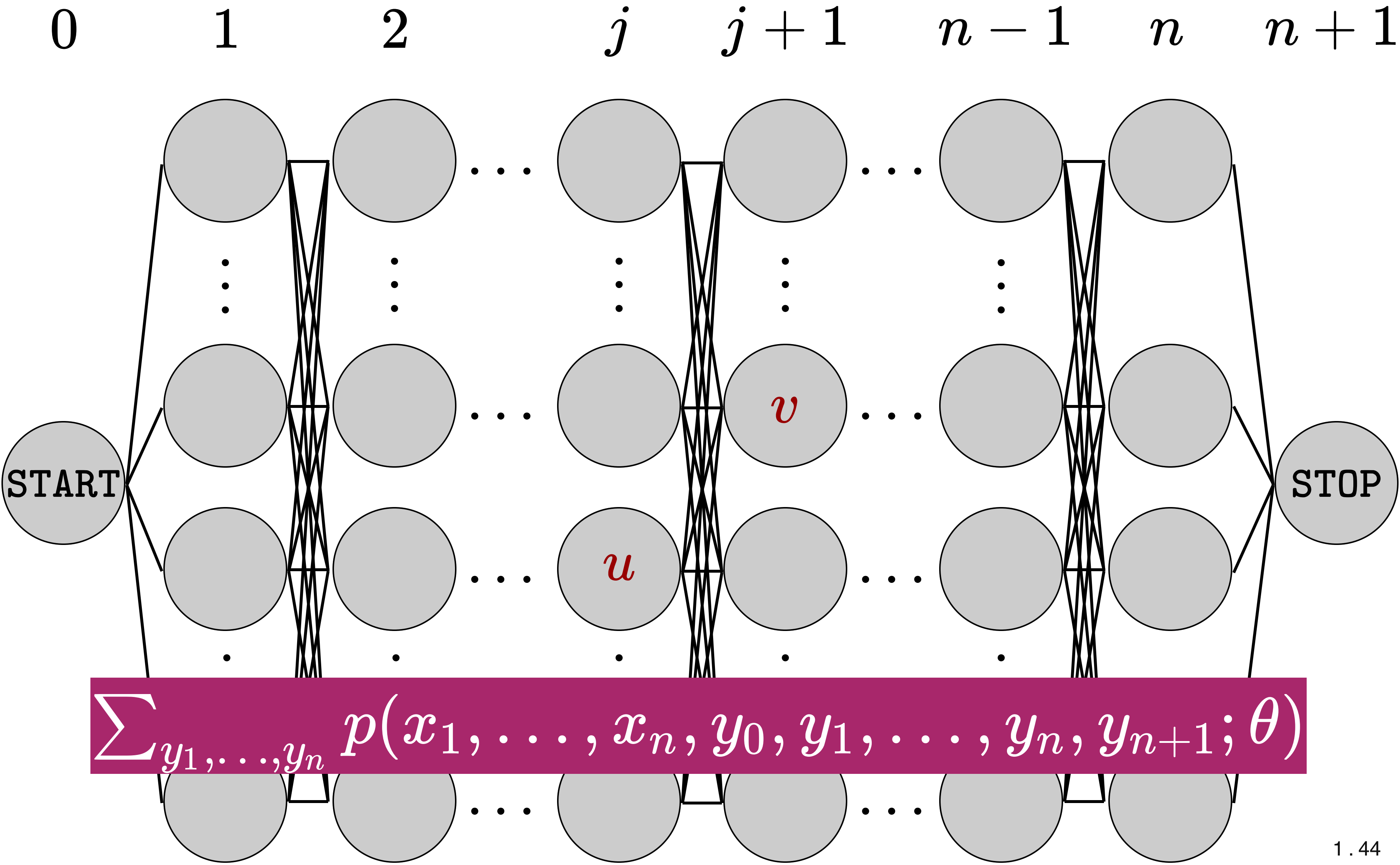
$$= \; p(x_1, \ldots, x_{j-1}, x_j, \ldots, x_n; \theta)$$

$$= \; \sum_{y_1,\ldots,y_n} p(x_1, \ldots, x_n, y_0, y_1, \ldots, y_n, y_{n+1}; \theta)$$

Decoding (Viterbi)

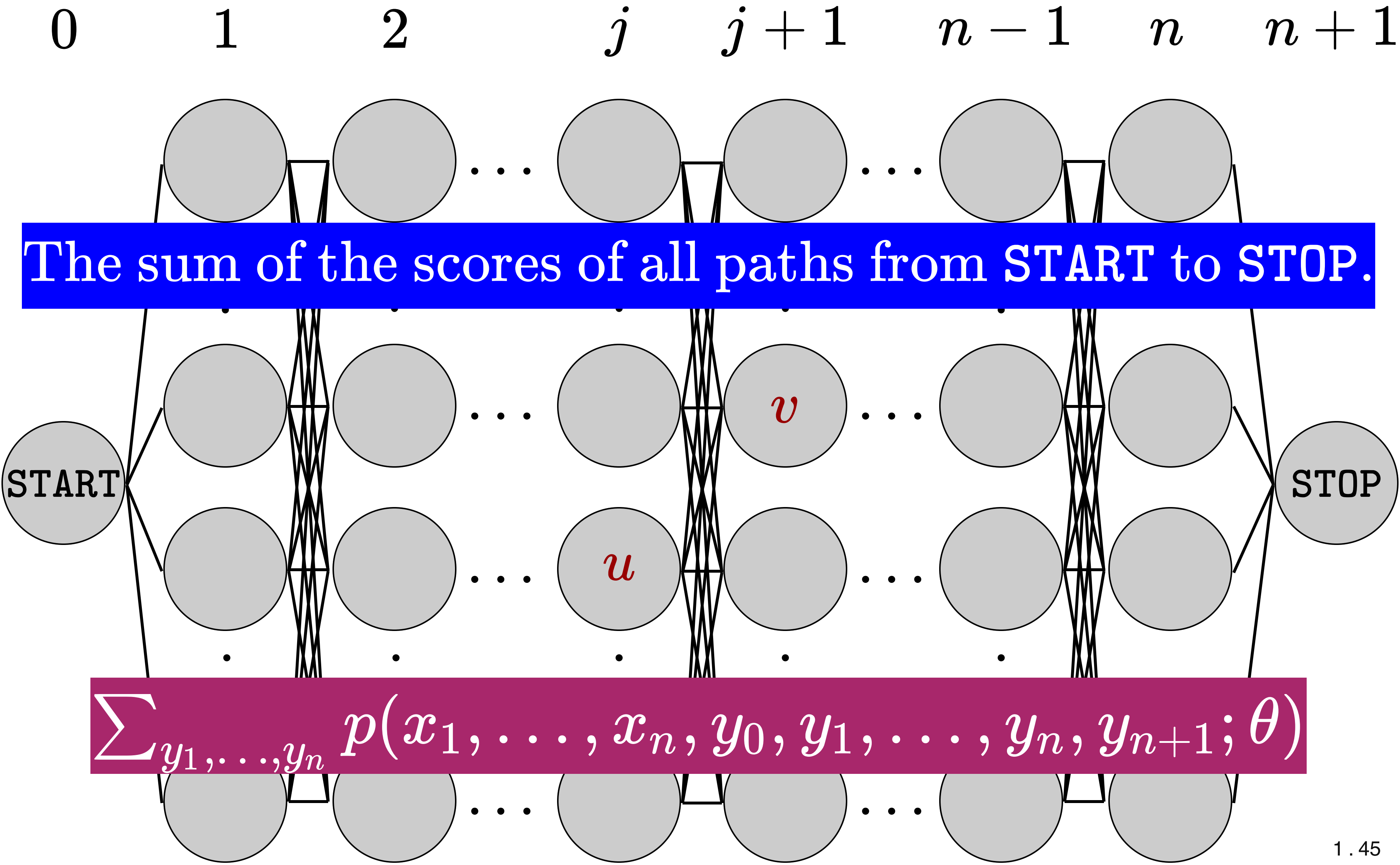
$$\max_{y_1,\ldots,y_n} p(x_1, \ldots, x_n, y_0, y_1, \ldots, y_n, y_{n+1}; \theta)$$

Inference in HMM



$$\sum_{y_1,\dots,y_n} p(x_1,\dots,x_n,y_0,y_1,\dots,y_n,y_{n+1};\theta)$$

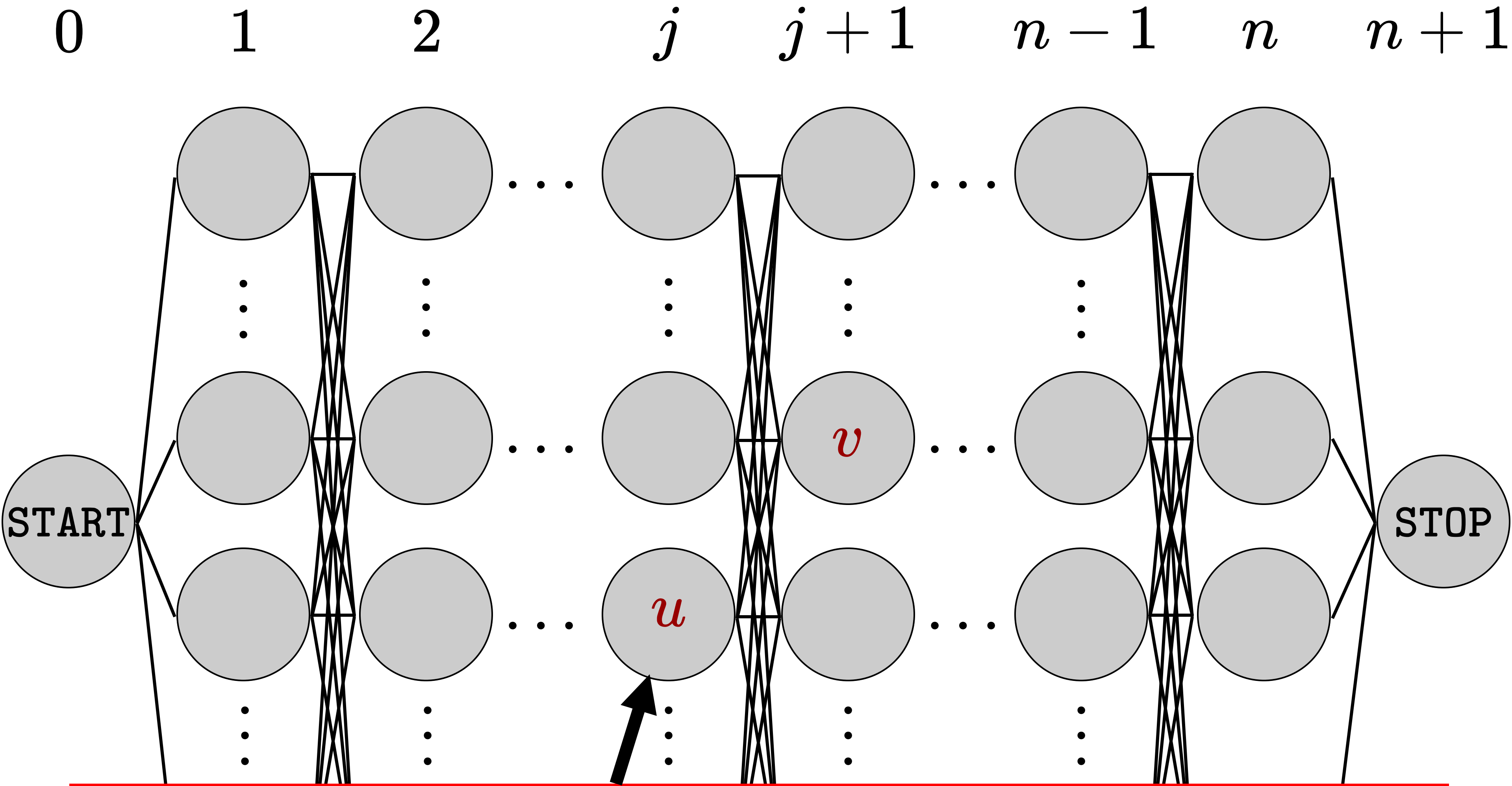
Inference in HMM



The sum of the scores of all paths from START to STOP.

$$\sum_{y_1,\ldots,y_n} p(x_1,\ldots,x_n,y_0,y_1,\ldots,y_n,y_{n+1};\theta)$$

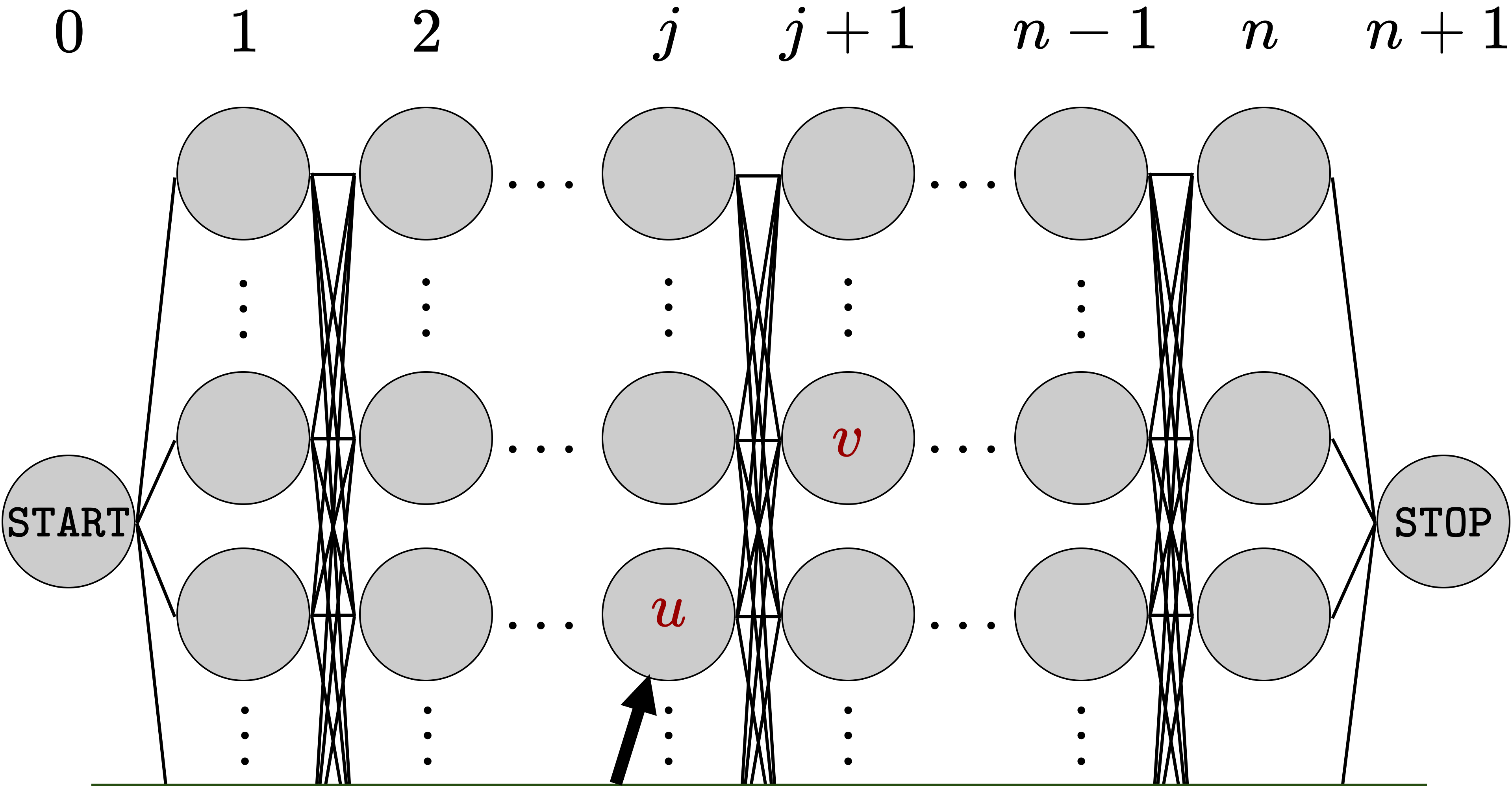
Forward-Backward Algorithm



$\alpha_u(j)$

The sum of the scores of all paths from **START** to this node.

Forward-Backward Algorithm



$\beta_u(j)$

The sum of the scores of all paths from this node to STOP

Forward-Backward Algorithm