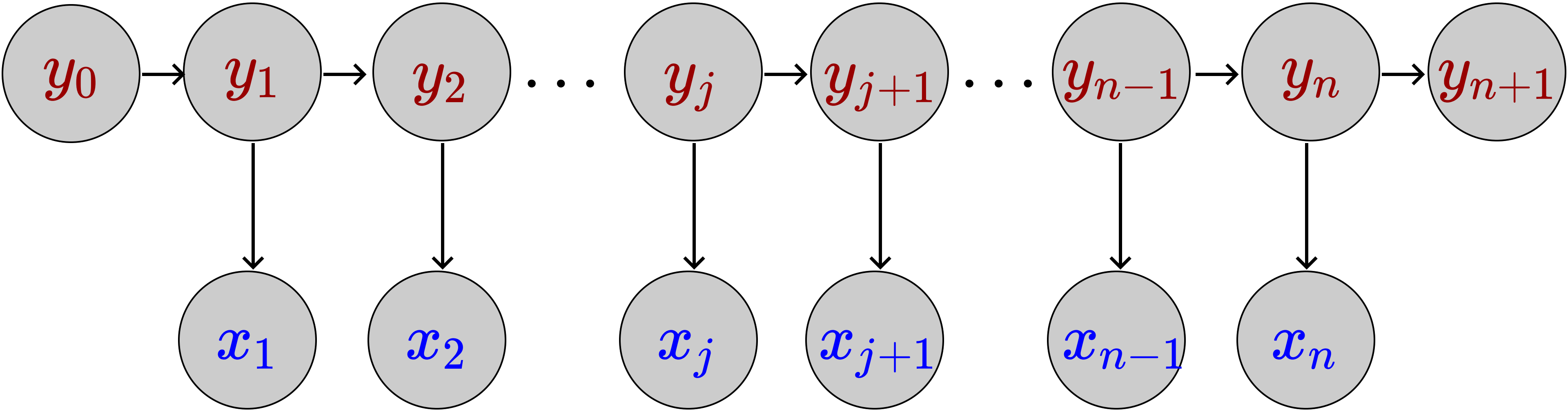
50.007
Machine Learning

Lu, Wei

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

Hidden Markov Model (II)

Hidden Markov Model
Parameterization

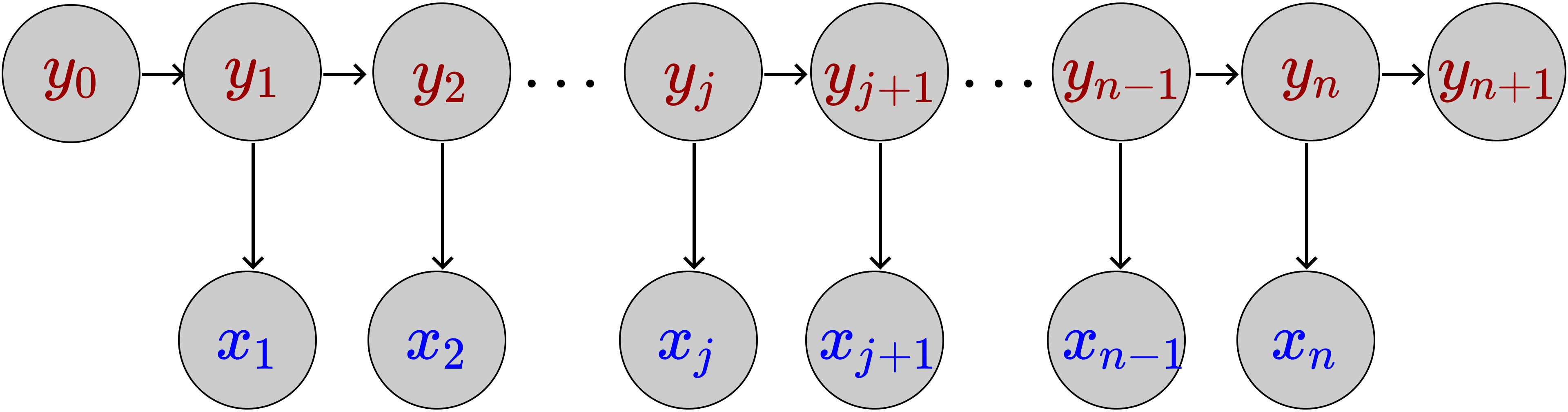


$$p(x_1, x_2, \ldots, x_{n-1}, x_n, y_0, y_1, y_2, \ldots, y_{n-1}, y_n, y_{n+1})$$

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

Hidden Markov Model

Supervised Learning



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

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

Hidden Markov Model

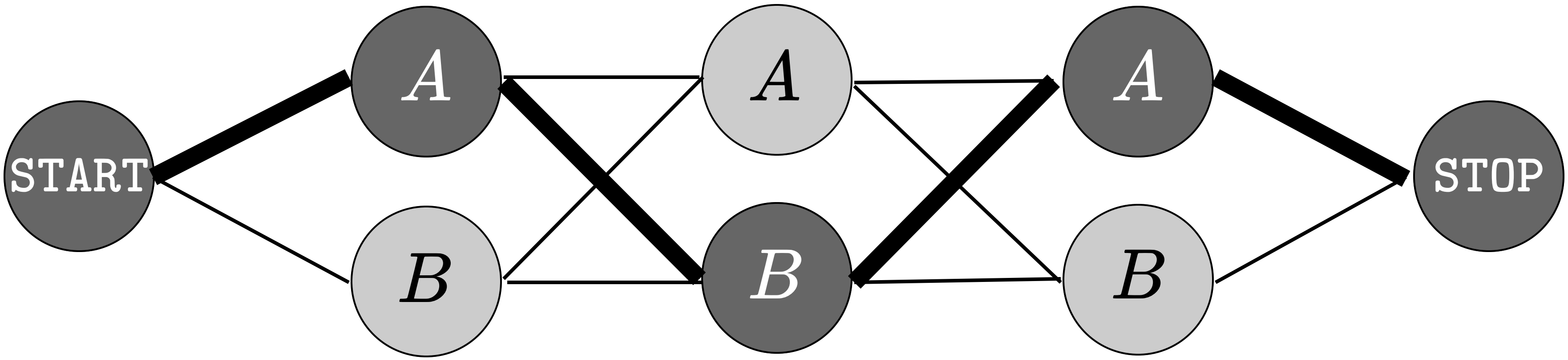
$$(\mathbf{x}, \mathbf{y}) = \text{the}/A, \text{dog}/B, \text{the}/A$$

$a_{u,v}$

$u \backslash v$	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



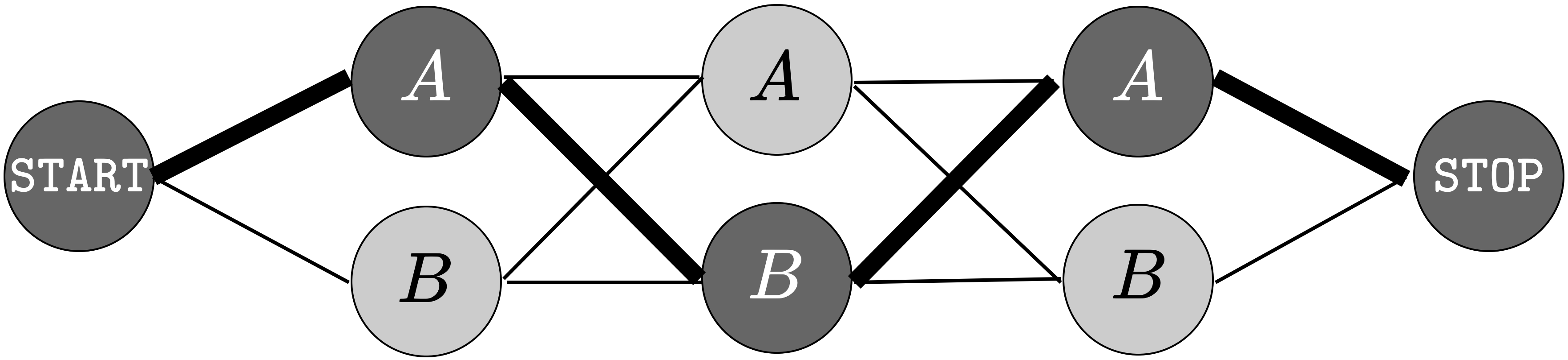
$$a_{\texttt{START},A} \times b_A(\text{“The”}) \times a_{A,B} \times b_B(\text{“dog”}) \times a_{B,A} \times b_A(\text{“the”}) \times a_{A,\texttt{STOP}}$$

Hidden Markov Model

$a_{u,v}$ $\quad (\mathbf{x}, \mathbf{y}) = \text{the}/B, \text{dog}/B, \text{the}/B \quad$ $b_u(o)$

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

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



What about this new $\mathbf{y}$ label sequence?

Hidden Markov Model

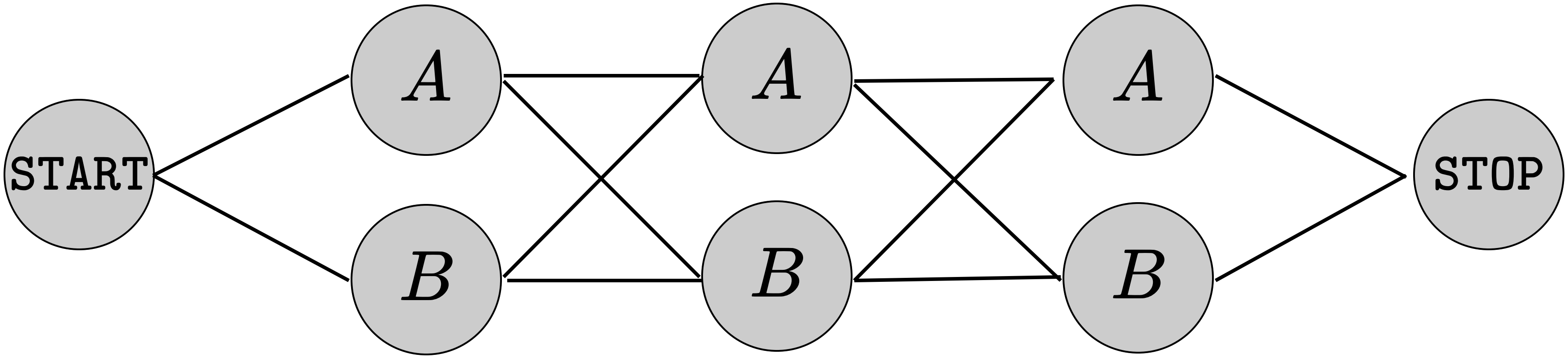
$a_{u,v}$

$u \backslash v$	A	B	STOP
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

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



Which label sequence $\mathbf{y}$ is the most probable given the word sequence $\mathbf{x}$?

Hidden Markov Model

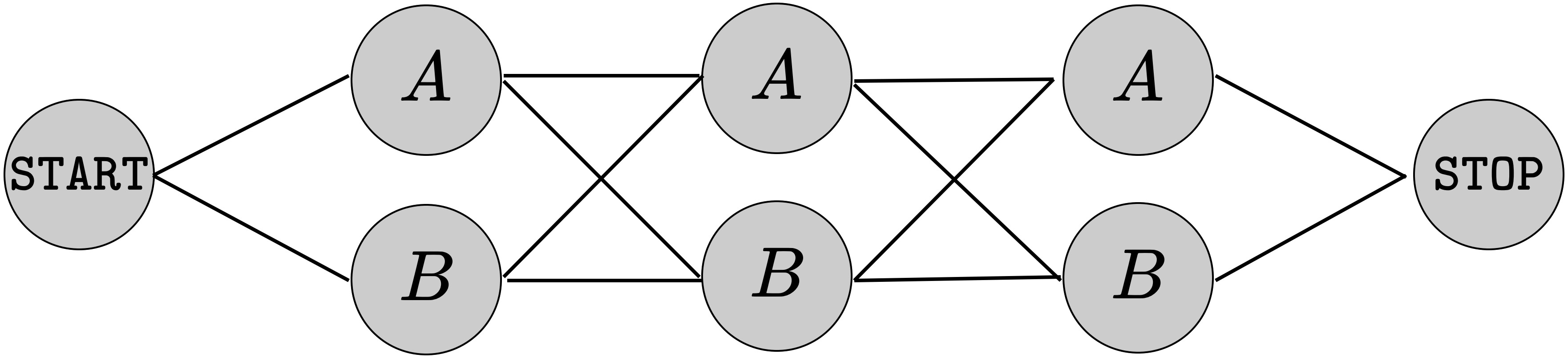
$a_{u,v}$

$u \backslash v$	A	B	STOP
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

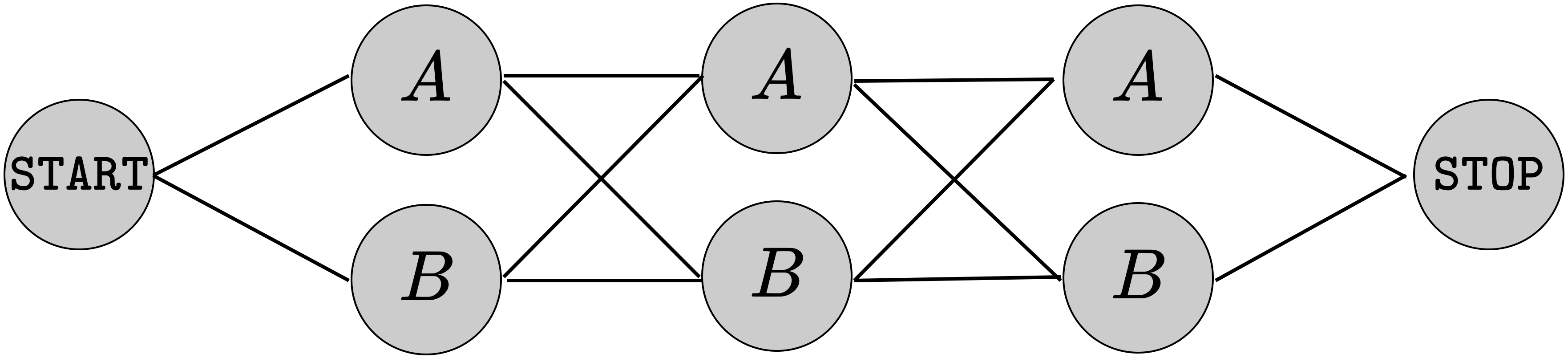
$$\mathbf{x} = \text{the dog the}$$



$$\mathbf{y}^* = \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})$$

Hidden Markov Model

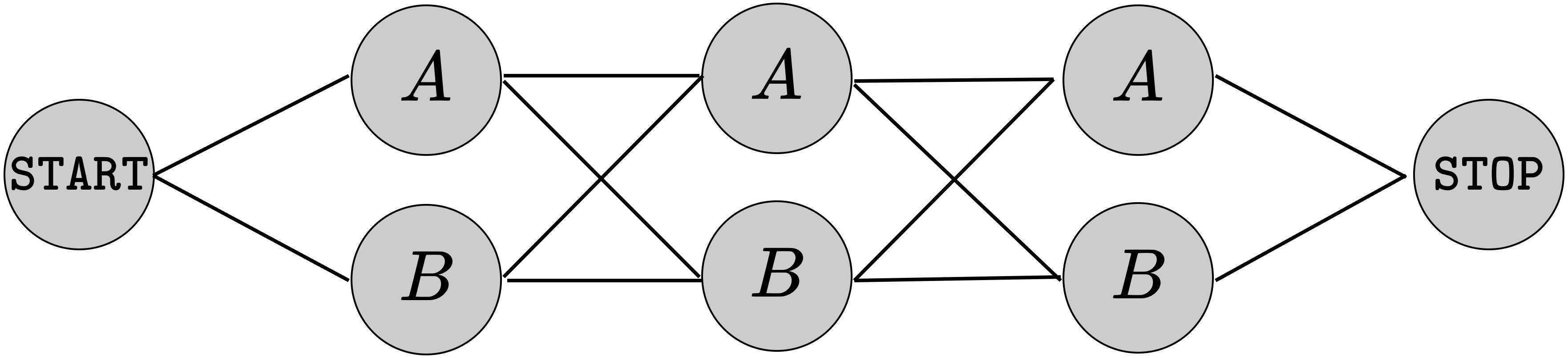
$$\mathbf{x} = \text{the dog the}$$



$$\mathbf{y}^* = \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})$$

Hidden Markov Model

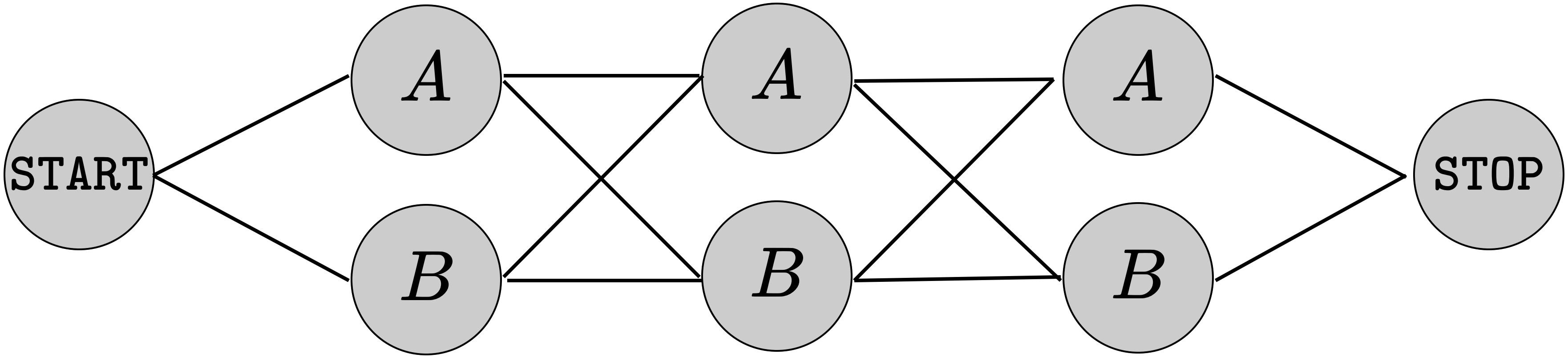
$$\mathbf{x} = \text{the dog the}$$



$$\begin{aligned}\mathbf{y}^* &= \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \\ &= \arg\max_{\mathbf{y}} p(\mathbf{x}, \mathbf{y})/p(\mathbf{x})\end{aligned}$$

Hidden Markov Model

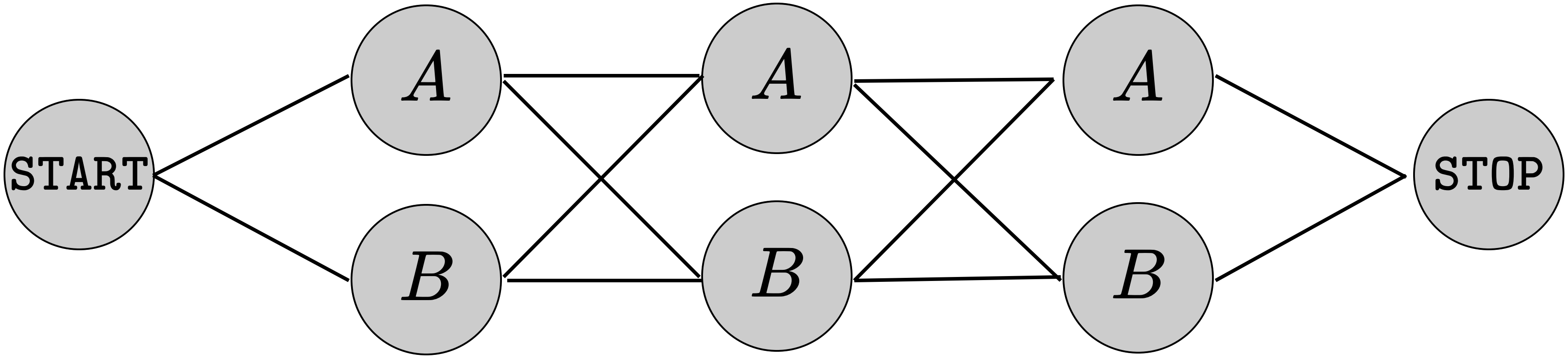
$$\mathbf{x} = \text{the dog the}$$



$$\begin{aligned}\mathbf{y}^* &= \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \\ &= \arg\max_{\mathbf{y}} p(\mathbf{x},\mathbf{y})/p(\mathbf{x}) \\ &= \arg\max_{\mathbf{y}} p(\mathbf{x},\mathbf{y})\end{aligned}$$

Hidden Markov Model

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

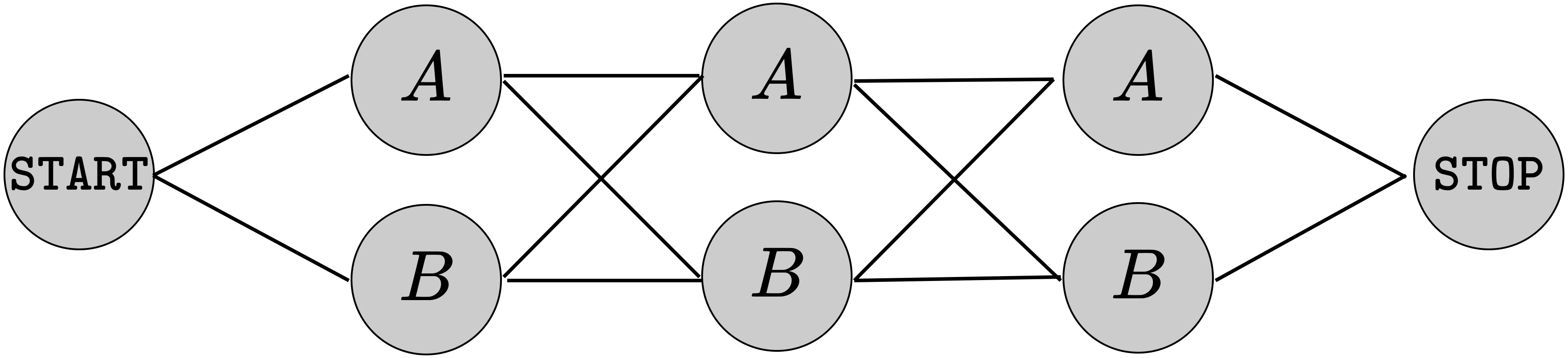


$$\begin{aligned}\mathbf{y}^* &= \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \\ &= \arg\max_{\mathbf{y}} p(\mathbf{x},\mathbf{y})/p(\mathbf{x}) \\ &= \arg\max_{\mathbf{y}} p(\mathbf{x},\mathbf{y})\end{aligned}$$

We can try one $\mathbf{y}$ at a time, and see which gives the highest score!

Hidden Markov Model

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



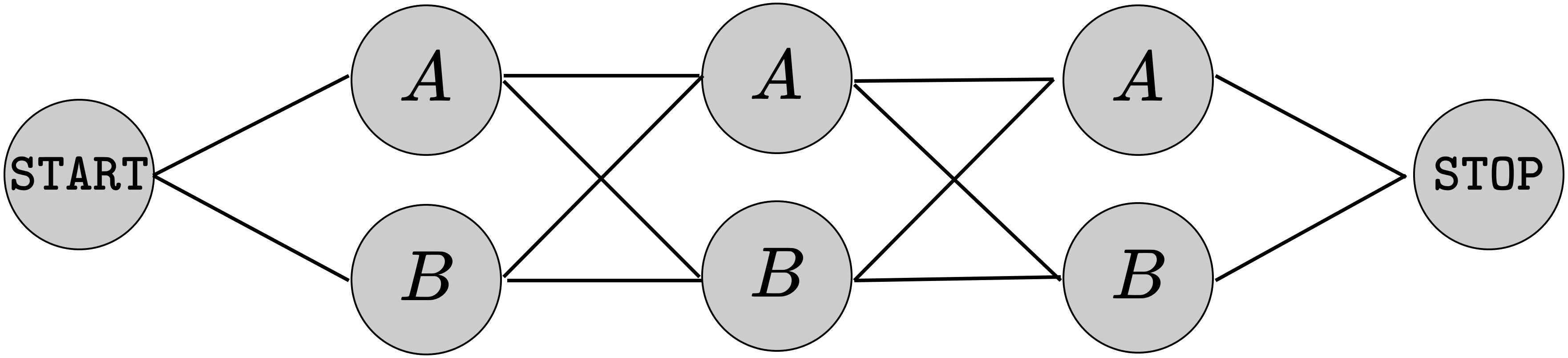
$$\begin{aligned}\mathbf{y}^* &= \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \\ &= \arg\max_{\mathbf{y}} p(\mathbf{x}, \mathbf{y})/p(\mathbf{x}) \\ &= \arg\max_{\mathbf{y}} p(\mathbf{x}, \mathbf{y})\end{aligned}$$

However, Is this a feasible approach in general?

How many possible label sequences to consider?

Hidden Markov Model

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



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

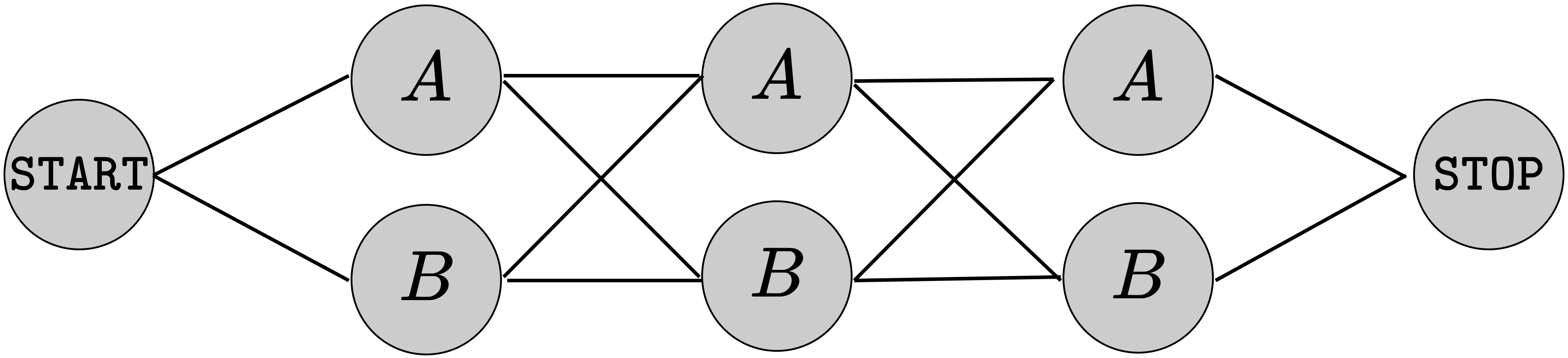
Number of words in the sentence

There are $O(|\mathcal{T}|^n)$ possible $\mathbf{y}$'s!

Number of possible tags at each word/position

Hidden Markov Model

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

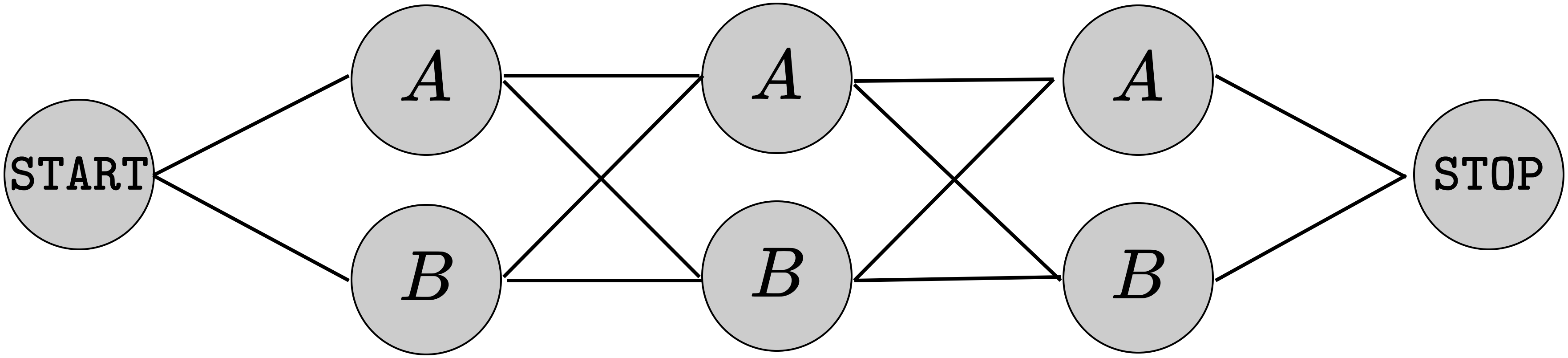


$$\mathbf{y}^* = \arg\max_{\mathbf{y}} p(\mathbf{x}, \mathbf{y})$$

We shall develop a more efficient approach for finding the most probable label sequence, or, the optimal path connecting START and STOP.

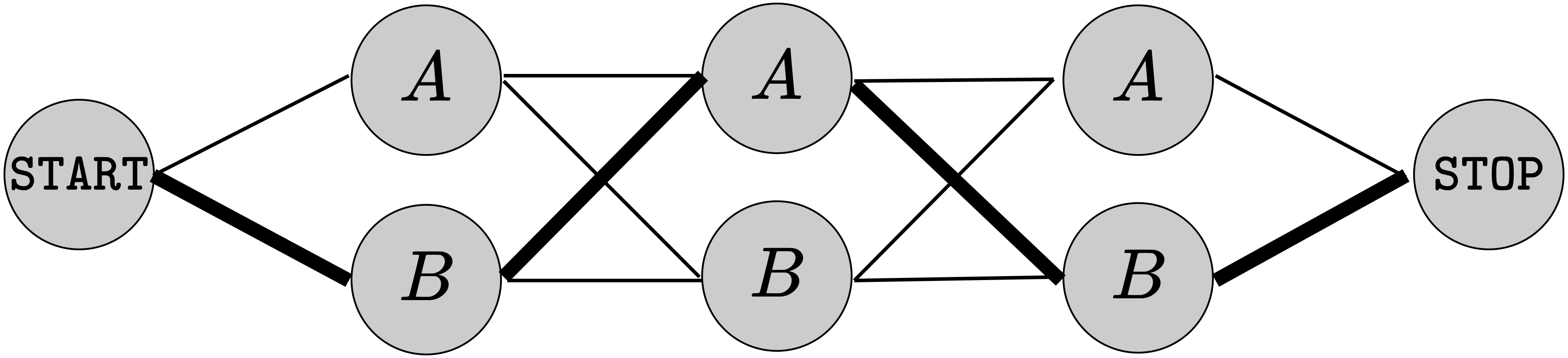
Hidden Markov Model

We are facing a problem of finding the highest scoring path connecting `START` and `STOP`



Hidden Markov Model

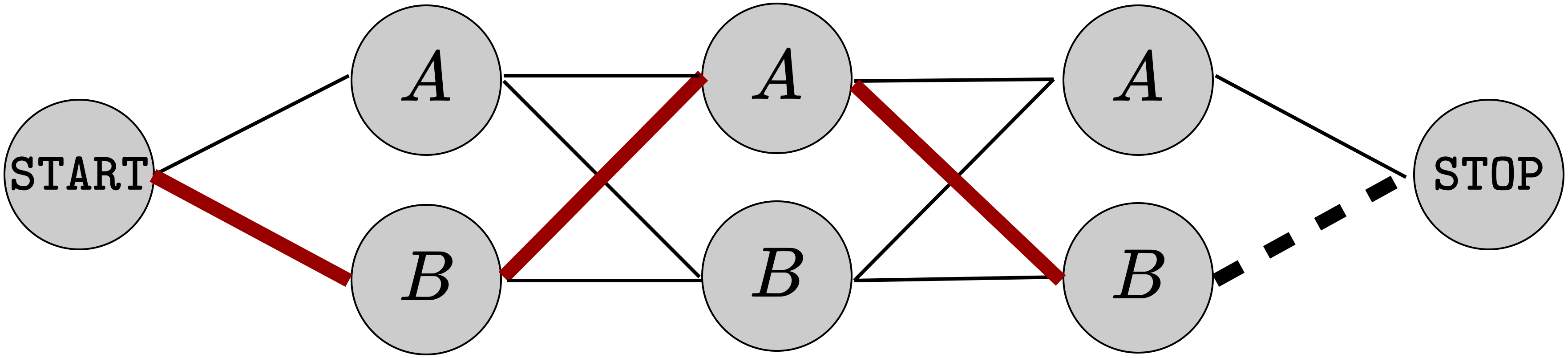
We are facing a problem of finding the highest scoring path connecting $\texttt{START}$ and $\texttt{STOP}$



Let's assume this is the highest scoring path

Hidden Markov Model

We are facing a problem of finding the highest scoring path connecting **START** and **STOP**

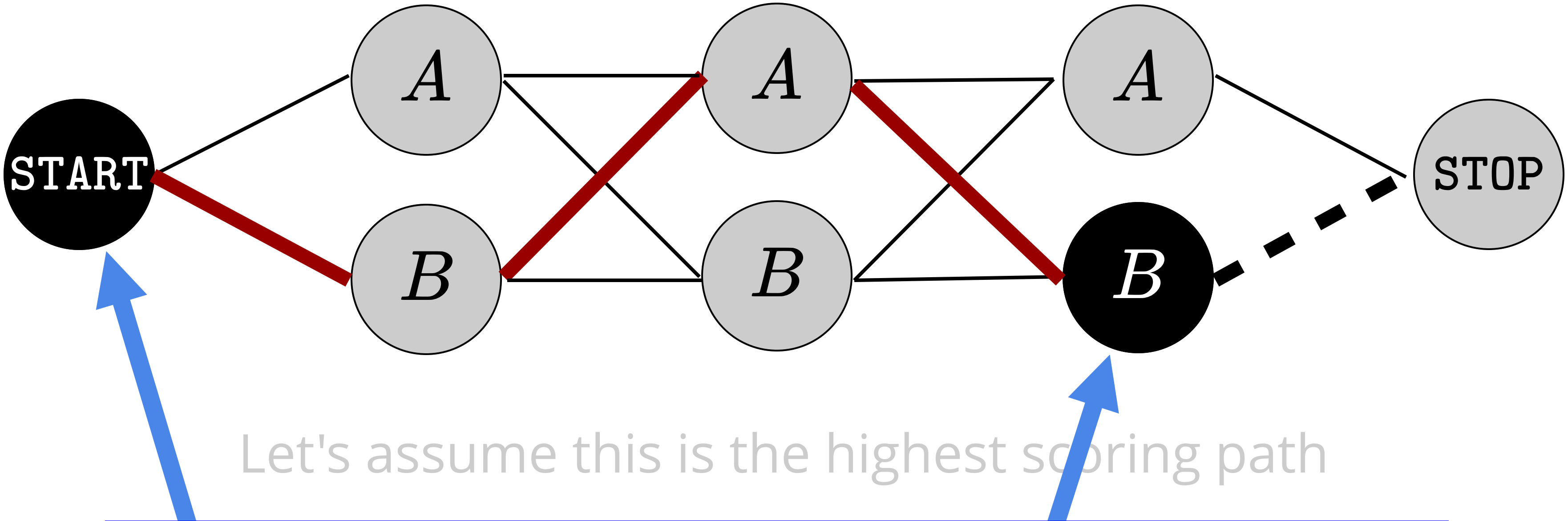


Let's assume this is the highest scoring path

Then, what can we say about this partial path?
(What types of properties do we know for this path?)

Hidden Markov Model

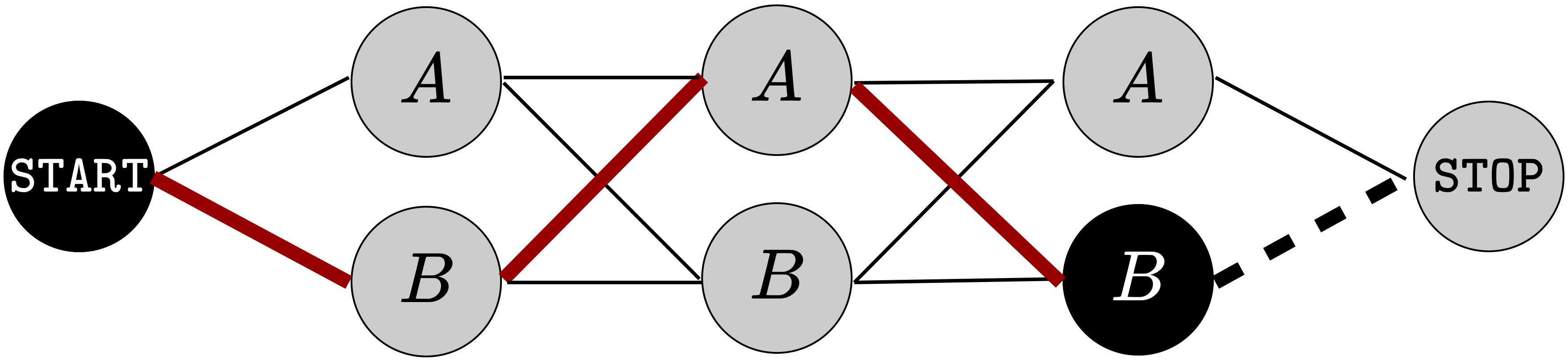
We are facing a problem of finding the highest scoring path connecting `START` and `STOP`



The highest scoring path among all paths connecting these two nodes!

Hidden Markov Model

We are facing a problem of finding the highest scoring path connecting **START** and **STOP**

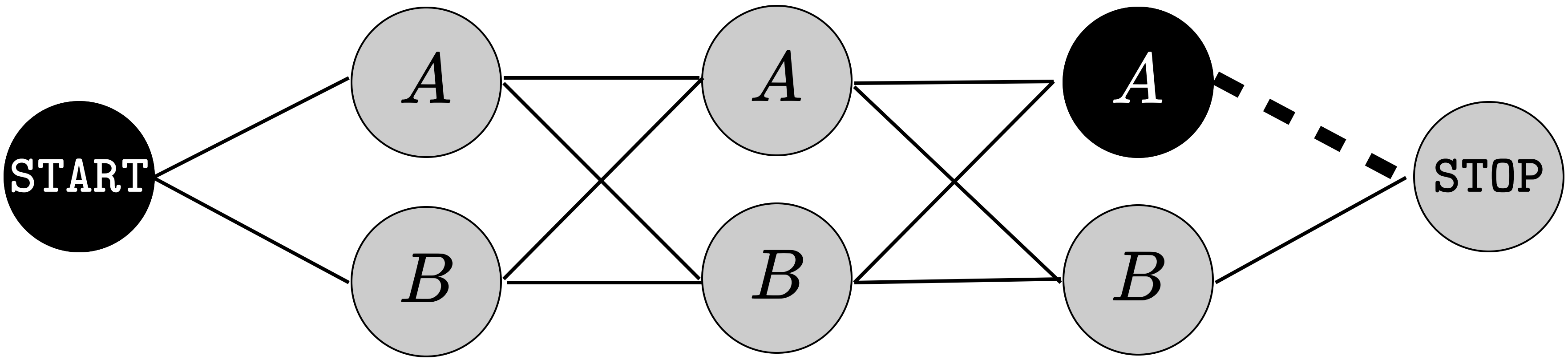


Is it possible to solve the original problem, if we already know the solutions to such sub-problems?

Hidden Markov Model

Case A

The second last node in the highest scoring path is A.



Find the highest scoring path from **START** to A at position n

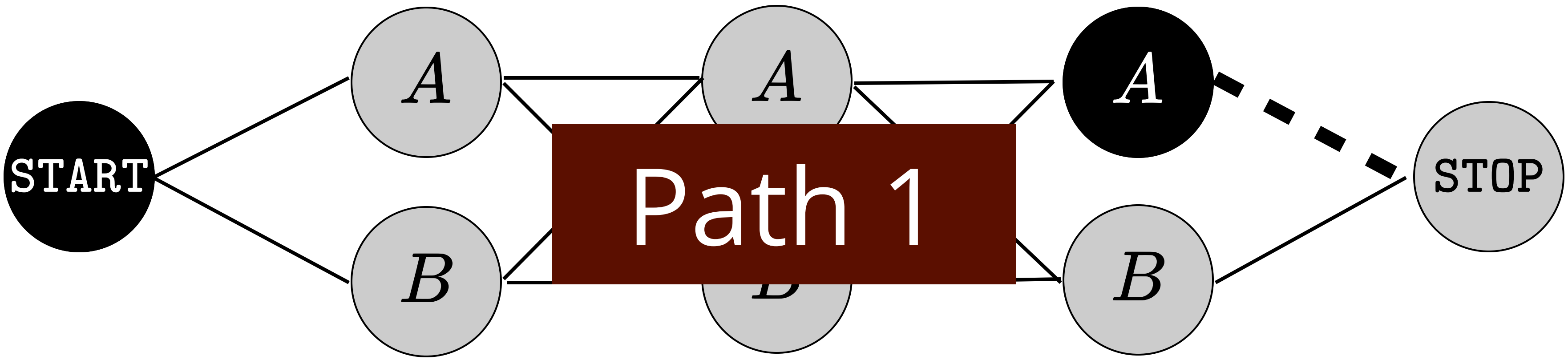
+

A single edge from from node A to **STOP**

Hidden Markov Model

Case A

The second last node in the highest scoring path is A.



Find the highest scoring path from START to A at position n

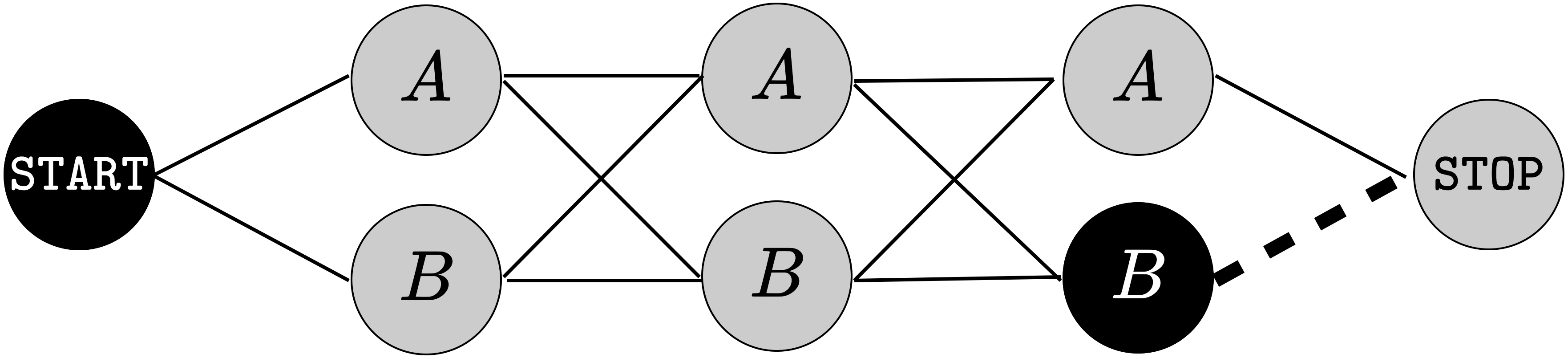
+

A single edge from from node A to STOP

Hidden Markov Model

Case B

The second last node in the highest scoring path is B.



Find the highest scoring path from $\texttt{START}$ to B at position n

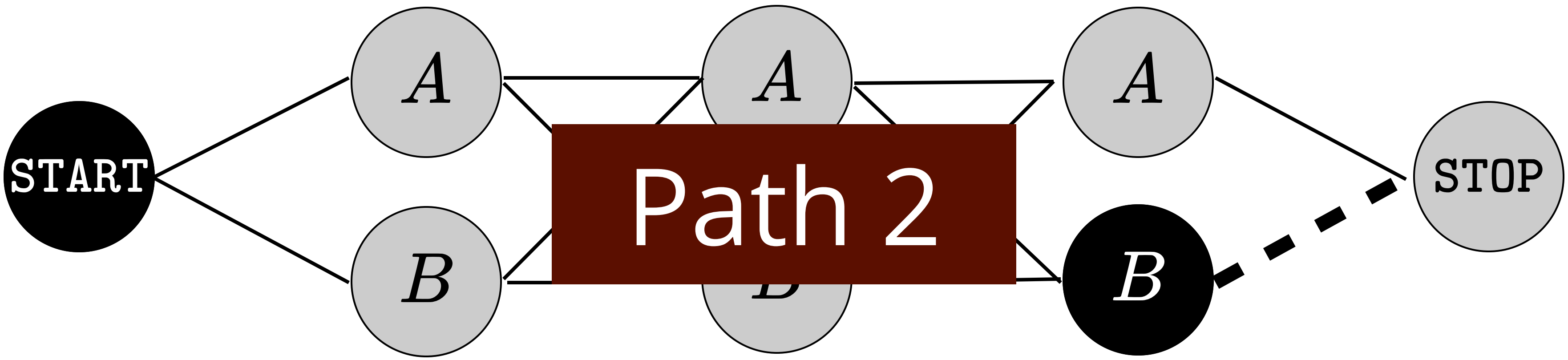
+

A single edge from from node B to $\texttt{STOP}$

Hidden Markov Model

Case B

The second last node in the highest scoring path is B.



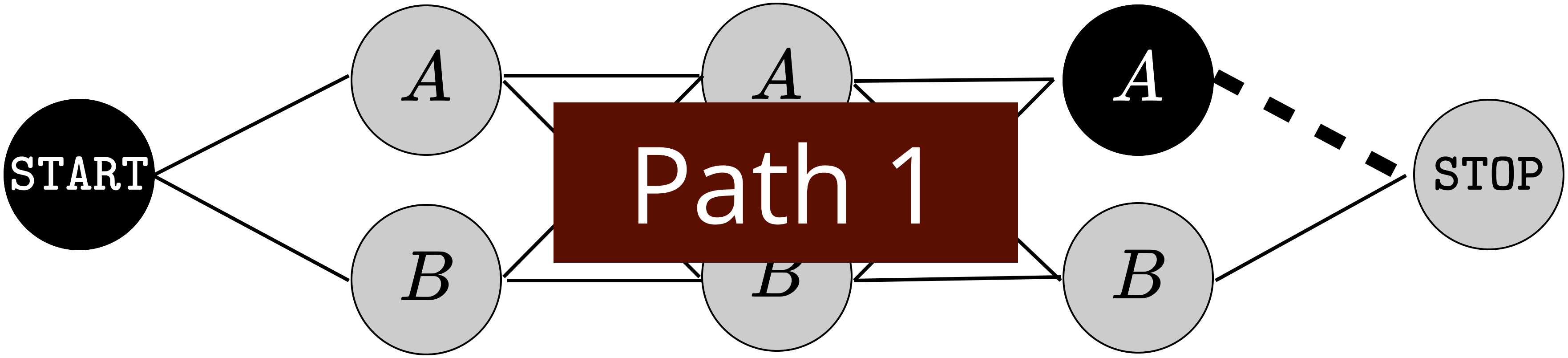
Find the highest scoring path from `START` to B at position n

+

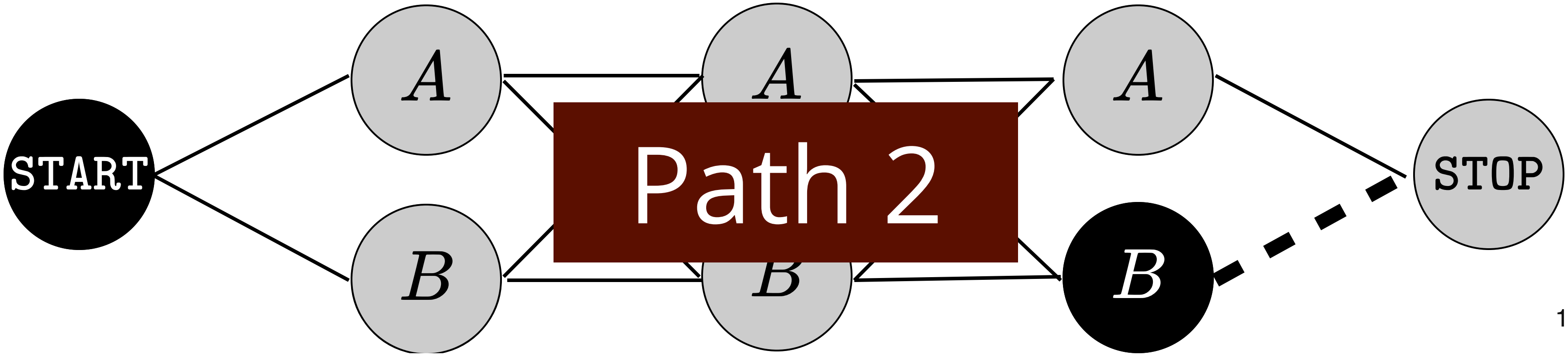
A single edge from from node B to `STOP`

Hidden Markov Model

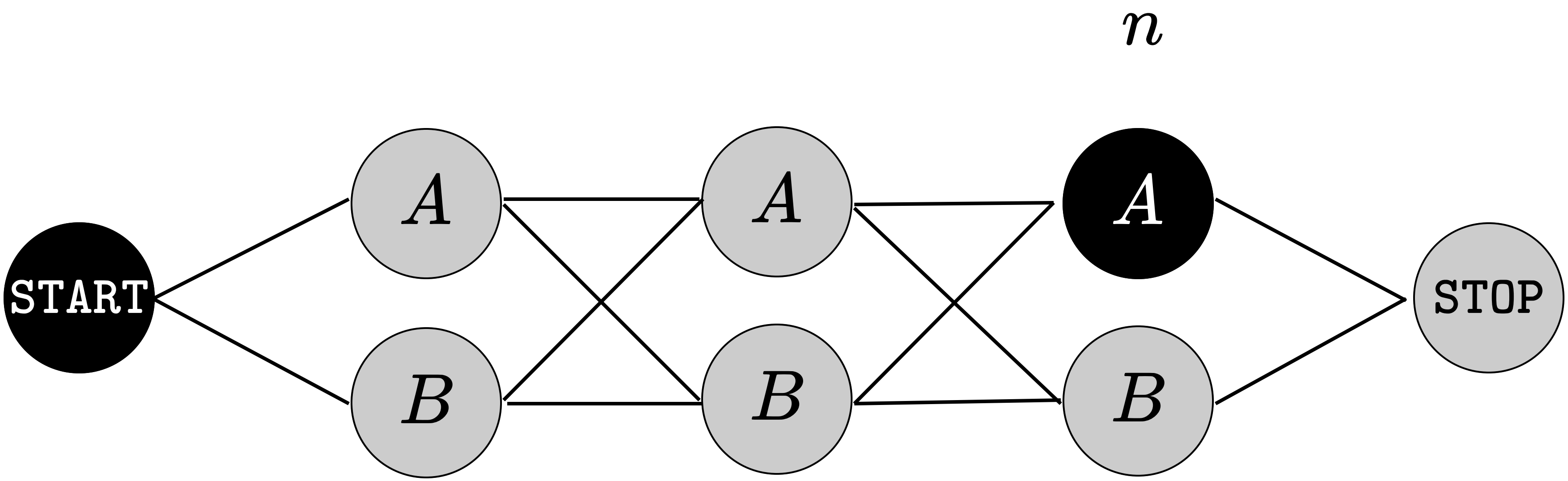
The highest scoring path from **START** to **STOP**



or



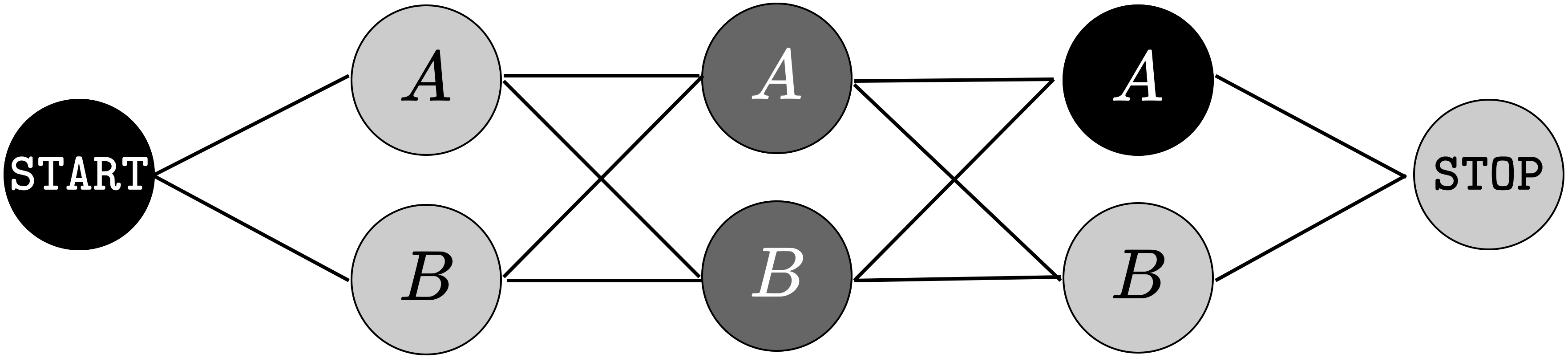
Hidden Markov Model



How do you find the highest scoring path from $\texttt{START}$ to node A at position n?

Hidden Markov Model

$$n-1$$



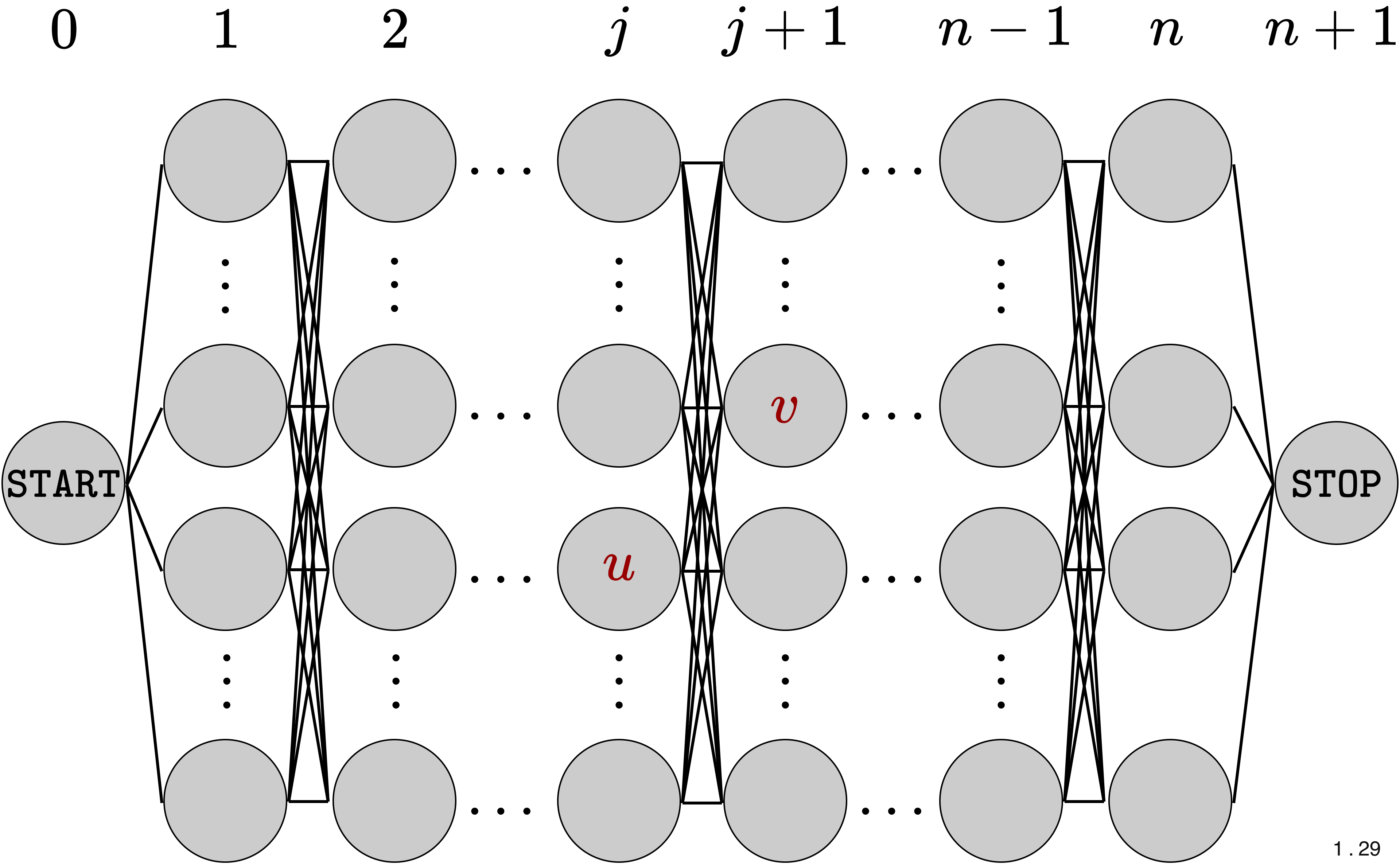
How do you find the highest scoring path from `START` to node A at position n?

We shall again rely on the partial paths from **`START`** to the two nodes at position $(n-1)$

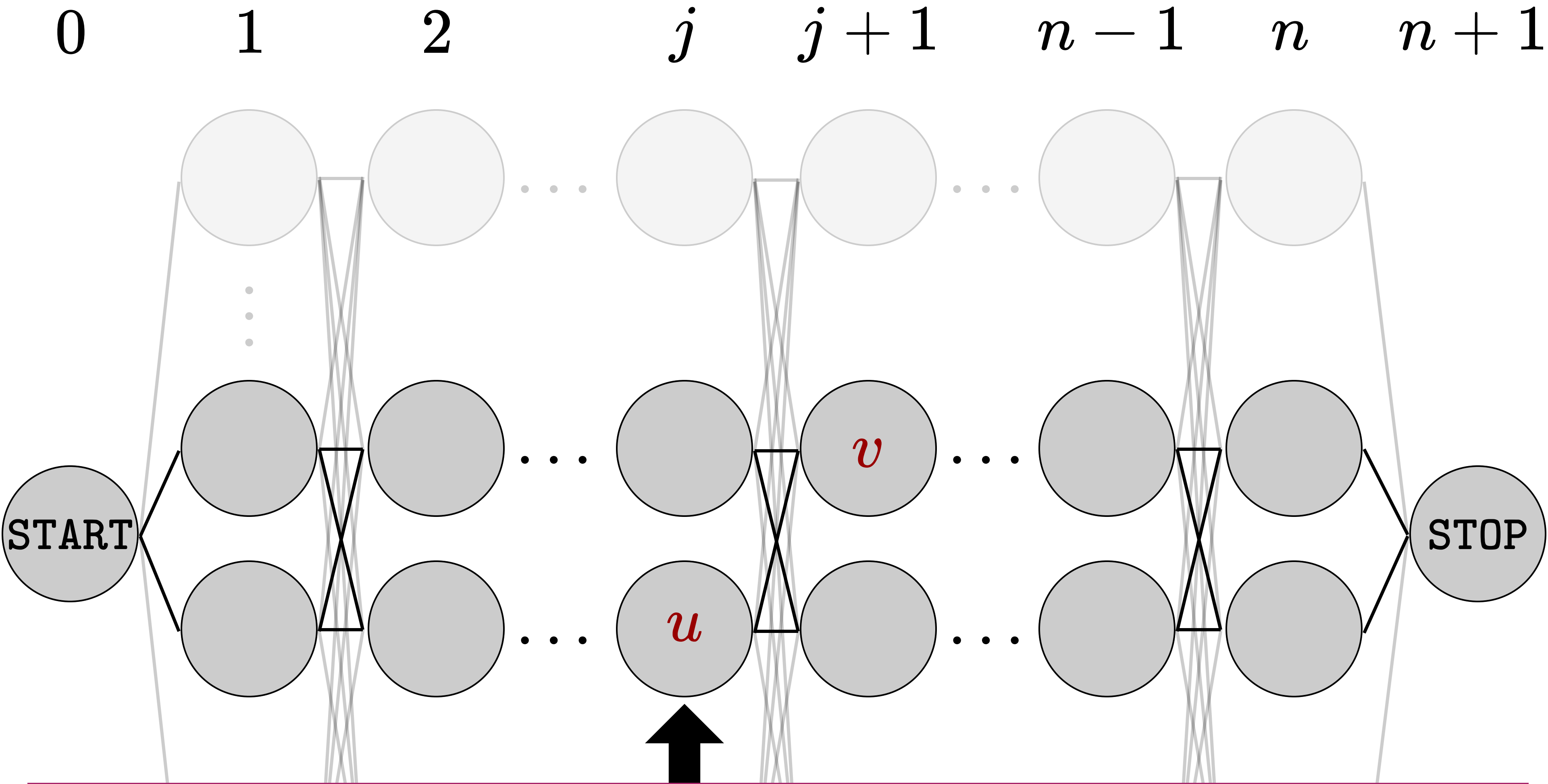
Question

Now, can we develop an efficient algorithm based on such observations?

Viterbi Algorithm

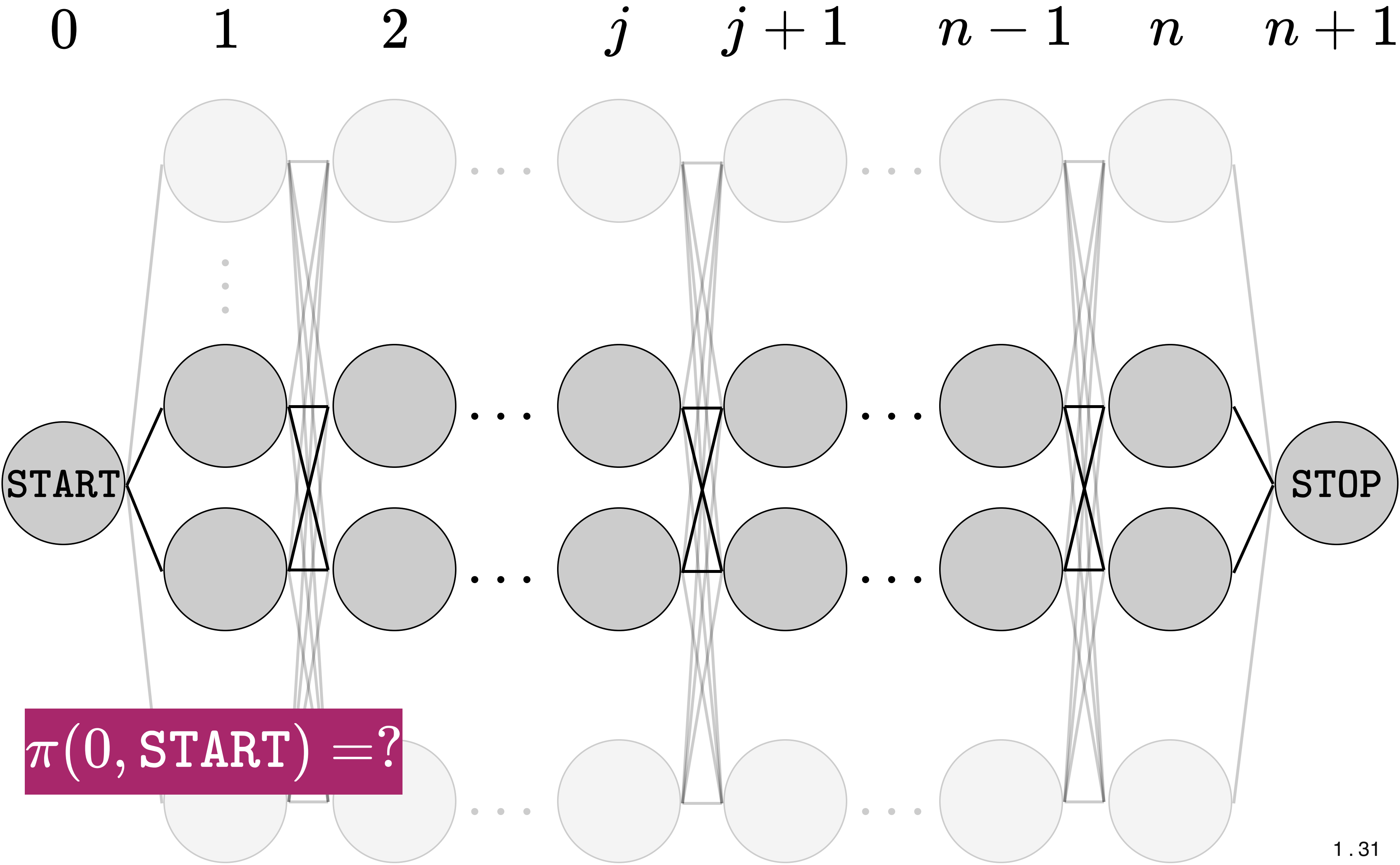


Viterbi Algorithm



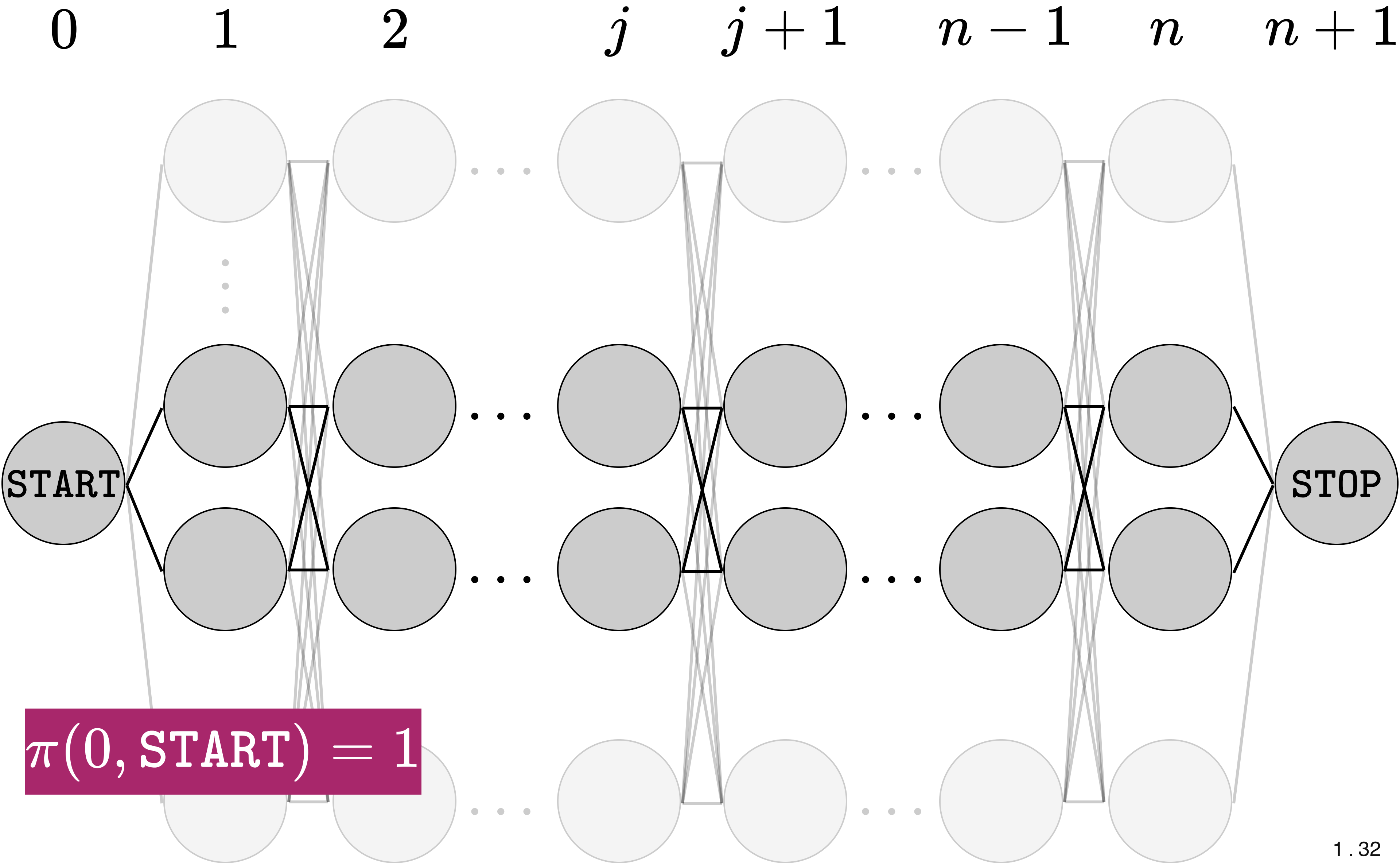
Store inside it $\pi(j, u)$ – the score of the highest scoring path from **START** to this node (j, u)

Viterbi Algorithm



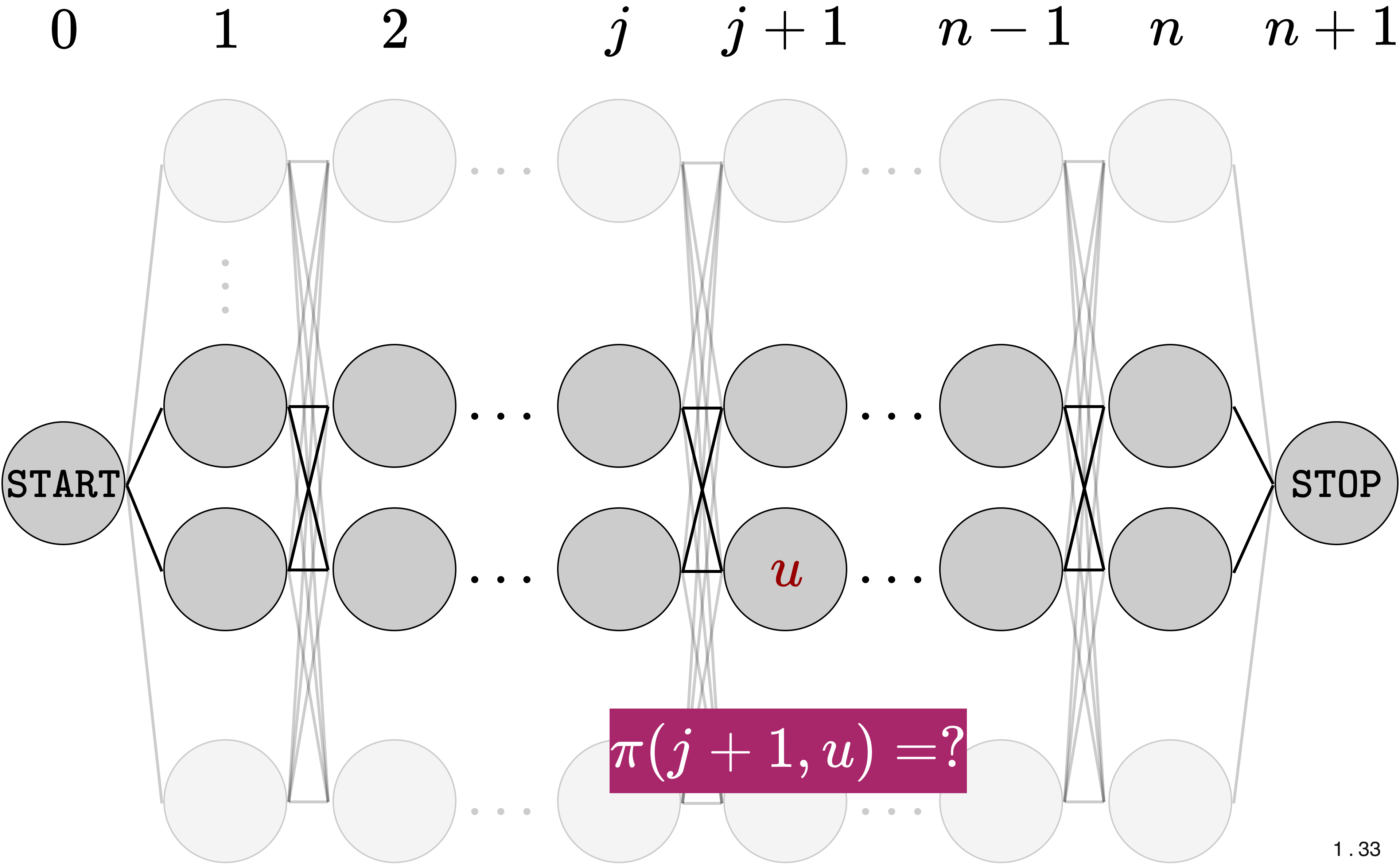
$\pi(0, \texttt{START}) = ?$

Viterbi Algorithm

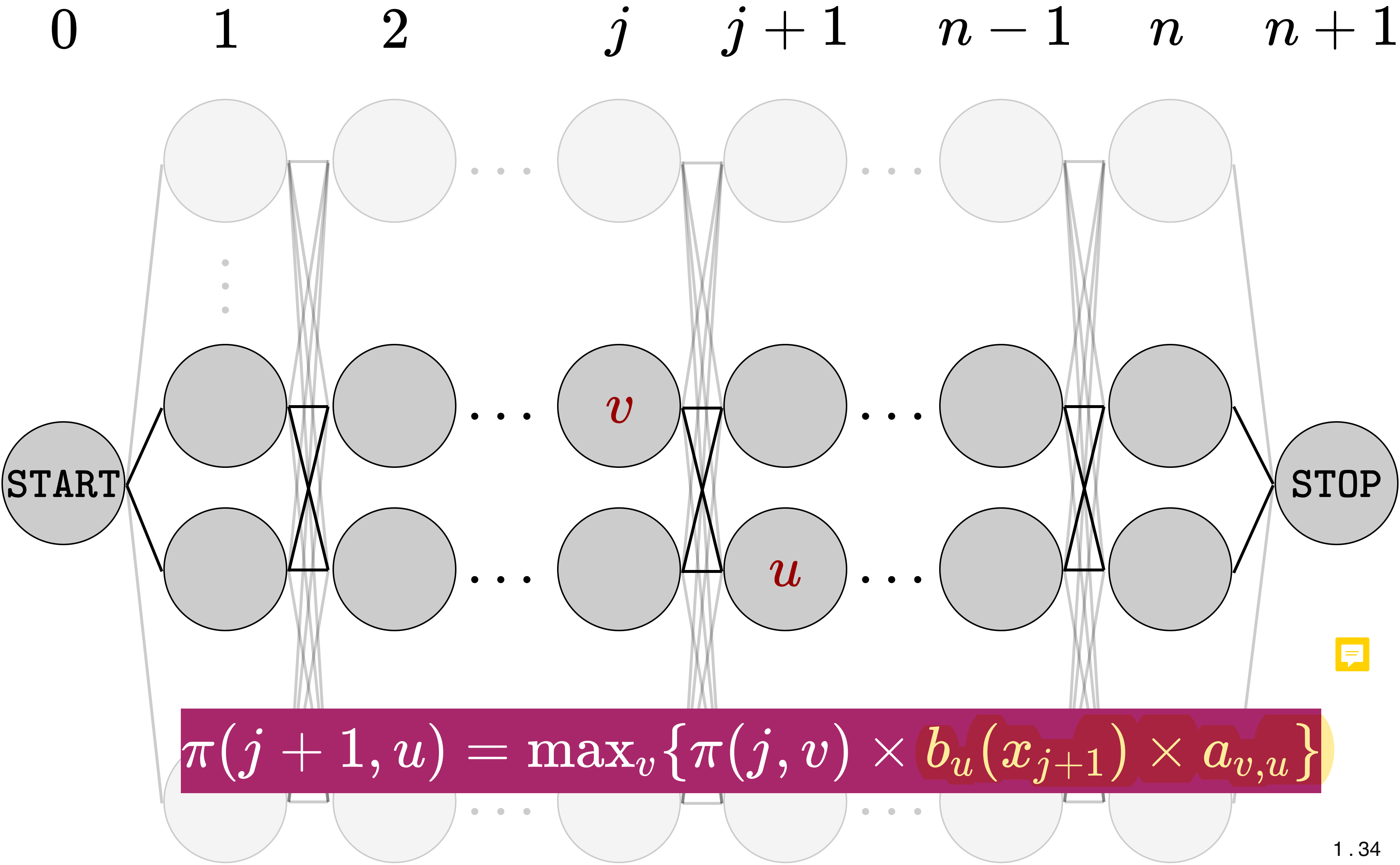


$\pi(0, \texttt{START}) = 1$

Viterbi Algorithm

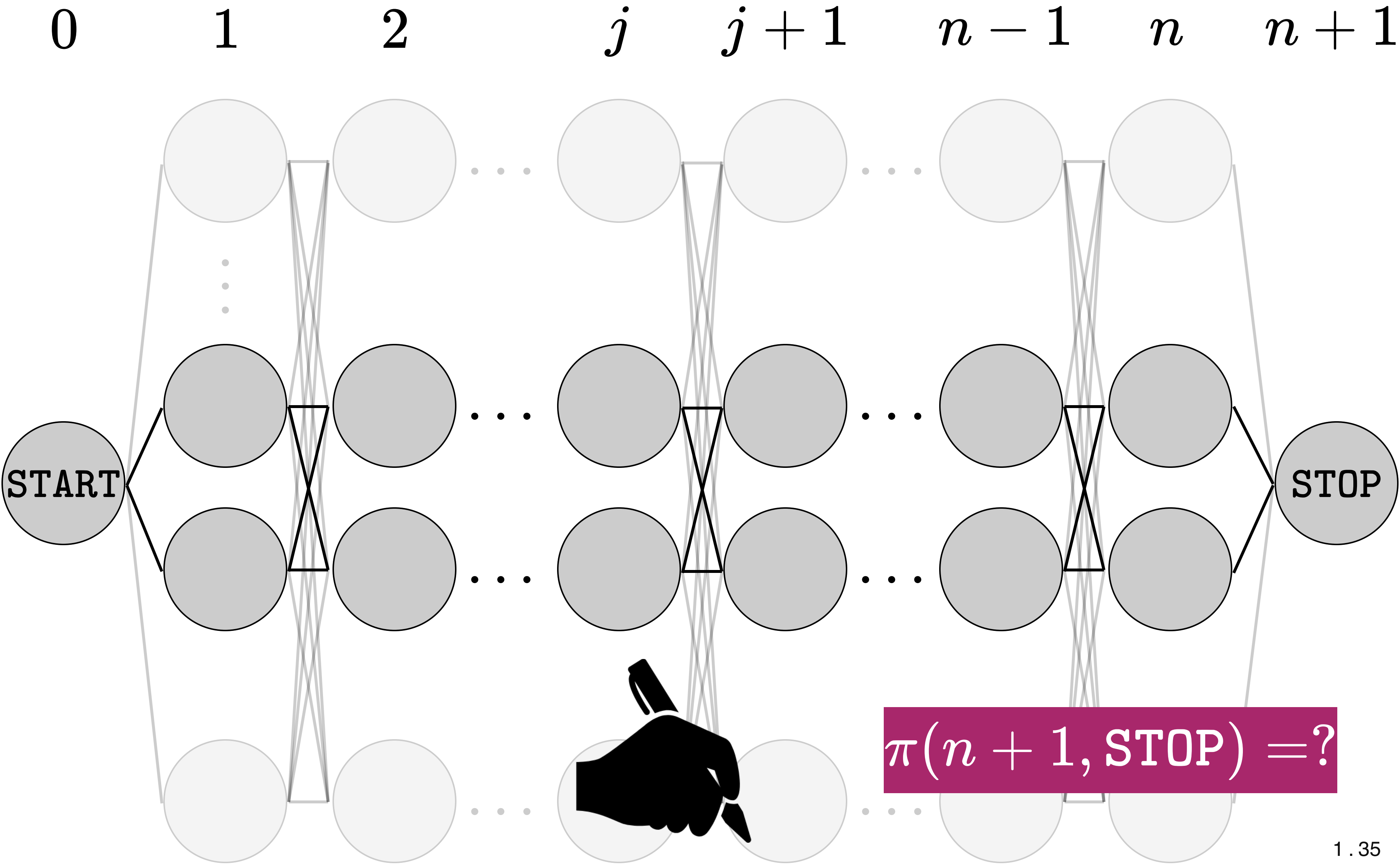


Viterbi Algorithm



$$\pi(j+1, u) = \max_v \{\pi(j, v) \times b_u(x_{j+1}) \times a_{v,u}\}$$

Viterbi Algorithm



Viterbi Algorithm

0 $\quad$ $n+1$

START $\quad$ STOP

1. Initialization

$$\pi(0,u) = \begin{cases} 1 & \text{if } u = \texttt{START} \\ 0 & \text{otherwise} \end{cases}$$

2. For $j = 0 \ldots n-1$, for each $u \in \mathcal{T}$

$$\pi(j+1,u) = \max_v \{\pi(j,v) \times b_u(x_{j+1}) \times a_{v,u}\}$$

3. Final Step

$$\pi(n+1,\texttt{STOP}) = \max_v \{\pi(n,v)) \times a_{v,\texttt{STOP}}\}$$

Viterbi Algorithm

0 $n+1$

START STOP

1. Initialization

$$\pi(0,u) = \begin{cases} 1 & \text{if } u = \texttt{START} \\ 0 & \text{otherwise} \end{cases}$$

2. For $j = 0 \dots n-1$, for each $u \in \mathcal{T}$

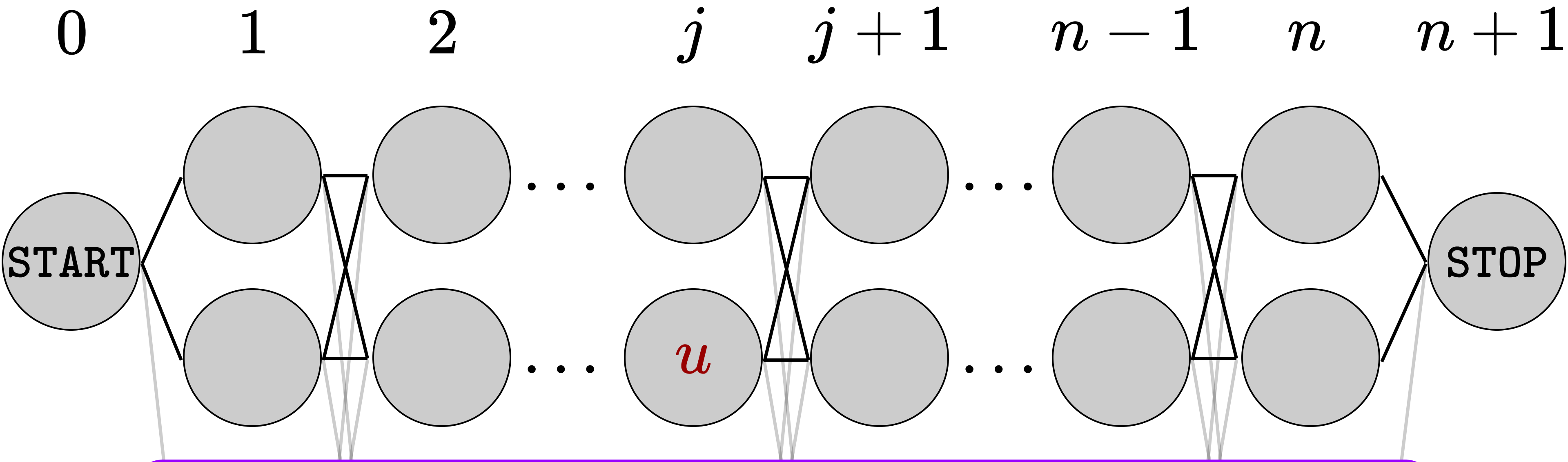
$$\pi(j+1,u) = \max_v \{\pi(j,v) \times b_u(x_{j+1}) \times a_{v,u}\}$$

3. Final Step

$$\pi(n+1,\texttt{STOP}) = \max_v \{\pi(n,v)) \times a_{v,\texttt{STOP}}\}$$

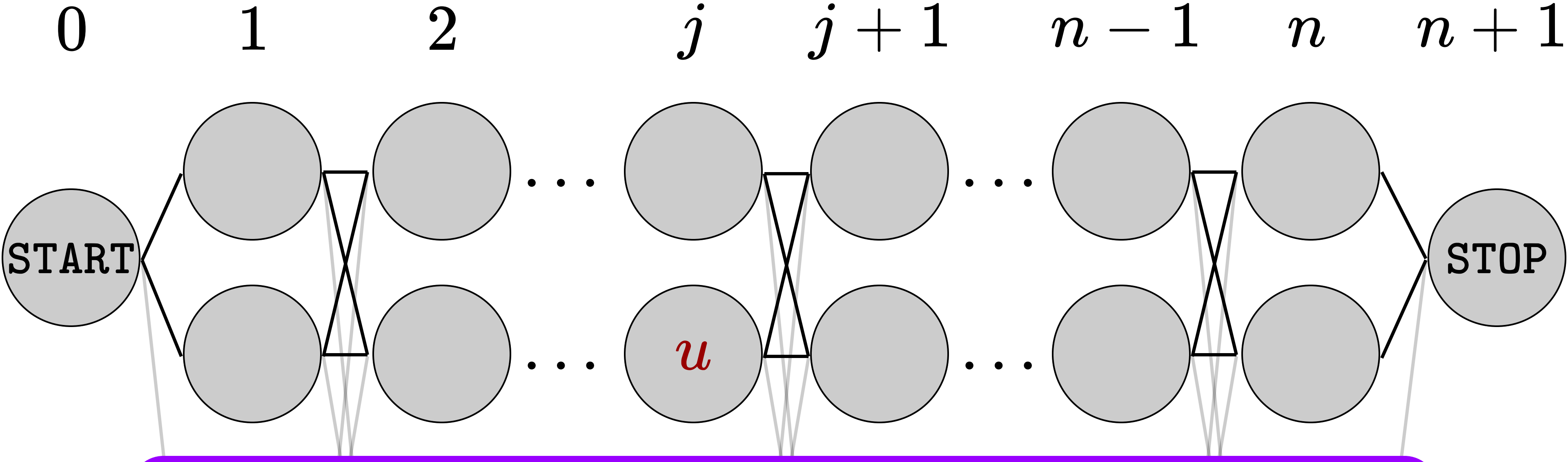
How do we figure out the highest scoring path from such scores?

Viterbi Algorithm



$$y_n^* = \arg\max_u \{\pi(n,u) \cdot a_{u,\texttt{STOP}}\}$$

Viterbi Algorithm

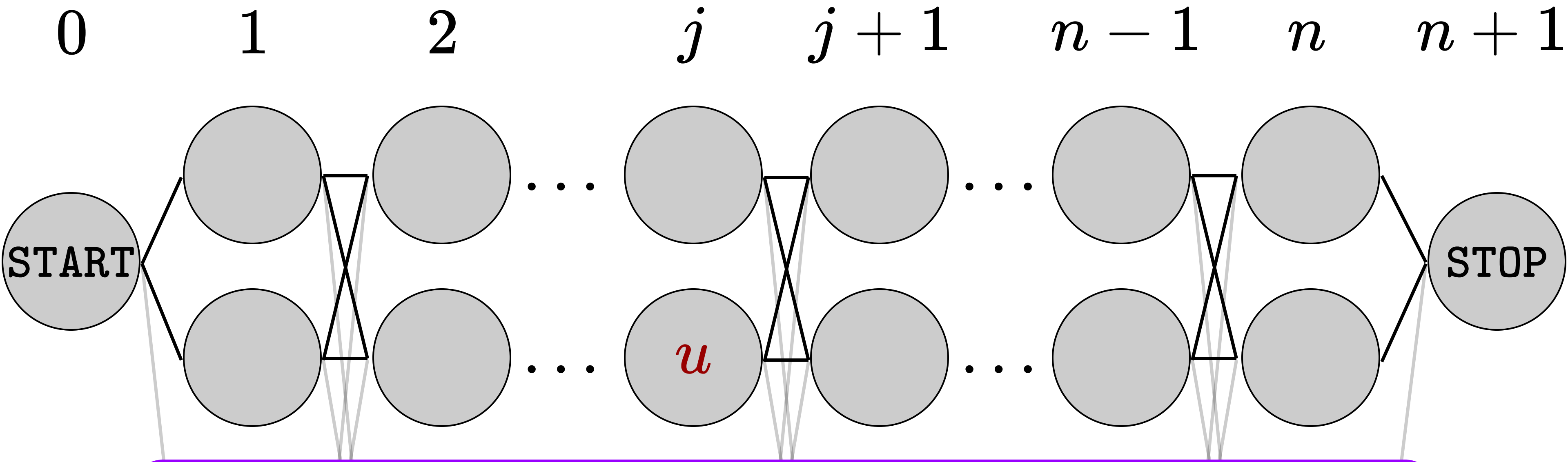


$$y_n^* = \arg\max_u \left\{\pi(n,u) \cdot a_{u,\texttt{STOP}}\right\}$$

For $j = n-1 \ldots 0$

$$y_j^* = \arg\max_u \{\pi(j,u) \cdot a_{u,y_{j+1}^*} \cdot b_{y_{j+1}^*}(x_{j+1})\}$$

Viterbi Algorithm



$$y_n^* = \arg\max_v \{\pi(n, v) \cdot a_{v,\texttt{STOP}}\}$$

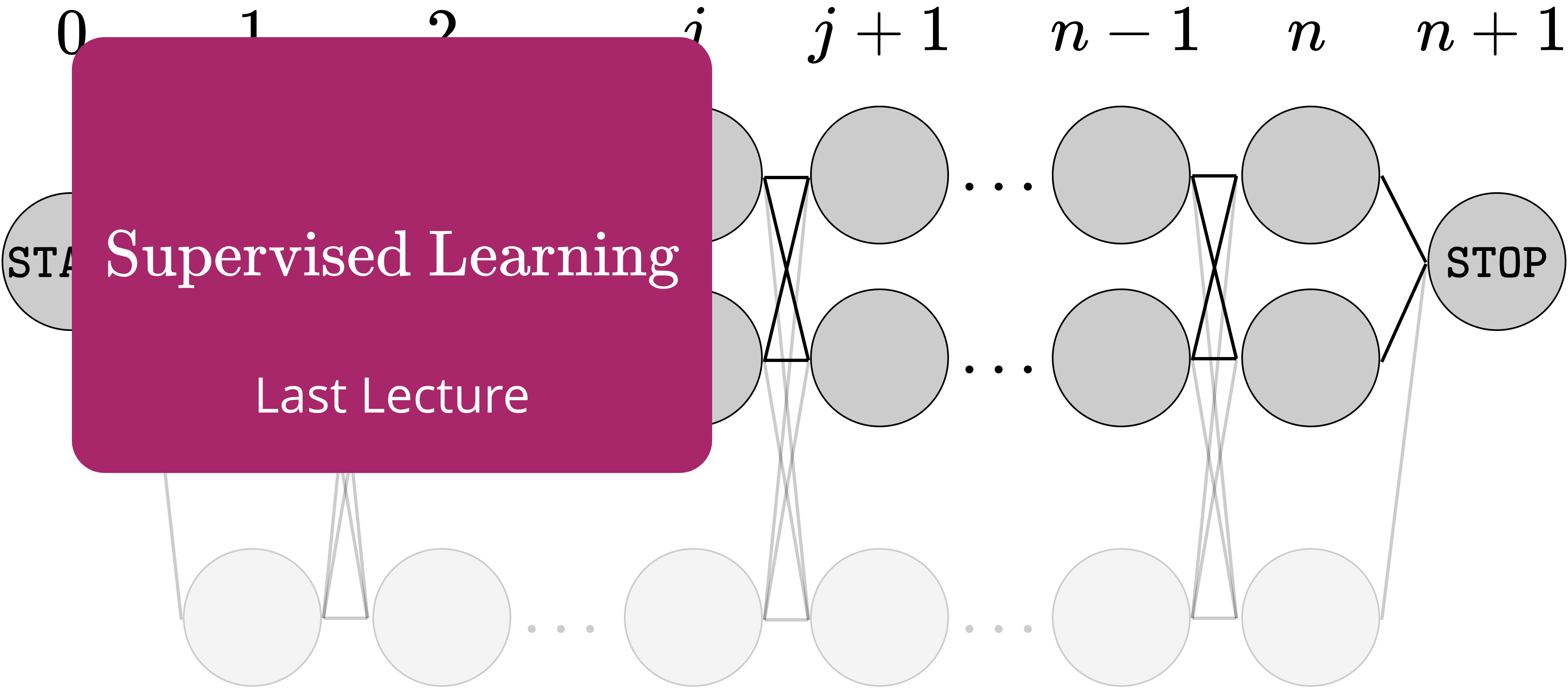
For $j = n - 1 \ldots 0$

$$y_j^* = \arg\max_u \{\pi(j, u) \cdot a_{u, y_{j+1}^*}\}$$

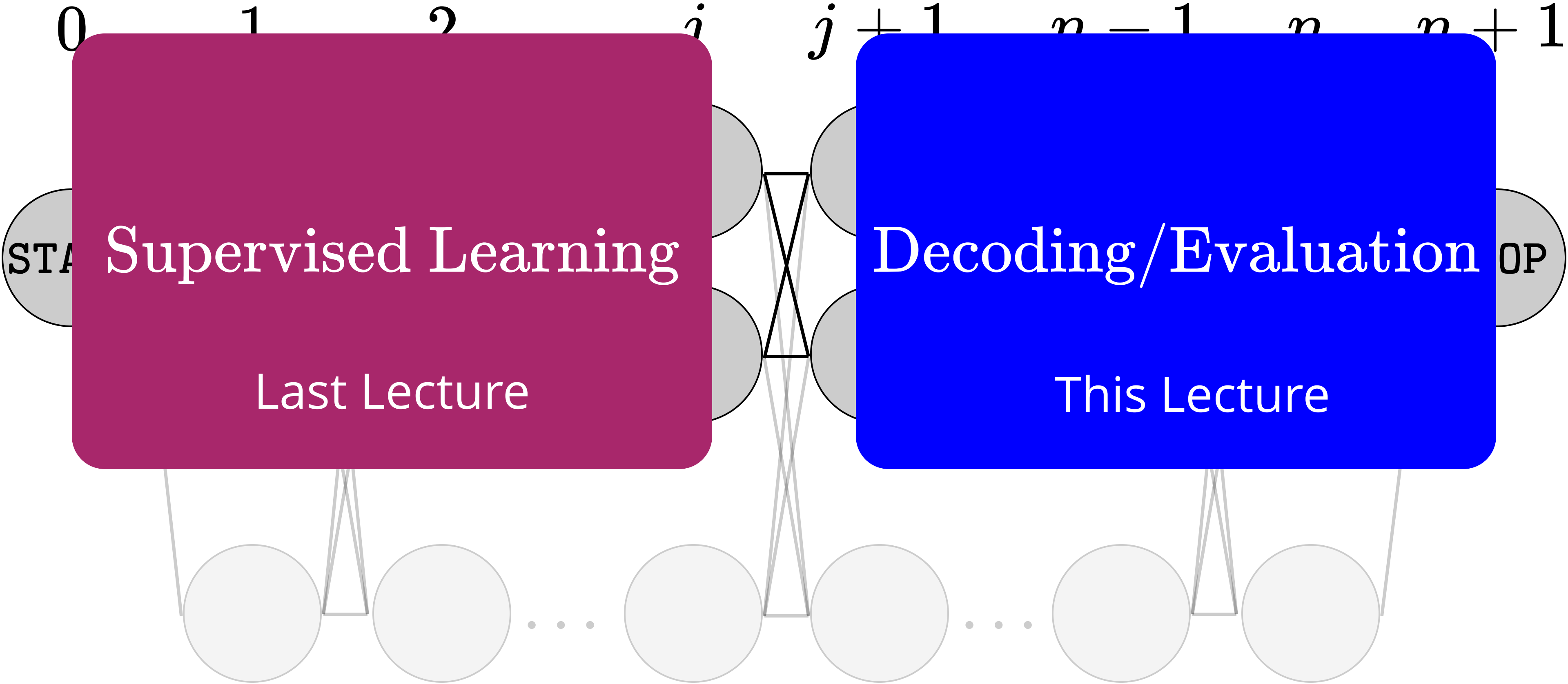
Question

What is the time complexity of the Viterbi algorithm?

Hidden Markov Model



Hidden Markov Model



Hidden Markov Model