50.007
Machine Learning

Lu, Wei

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

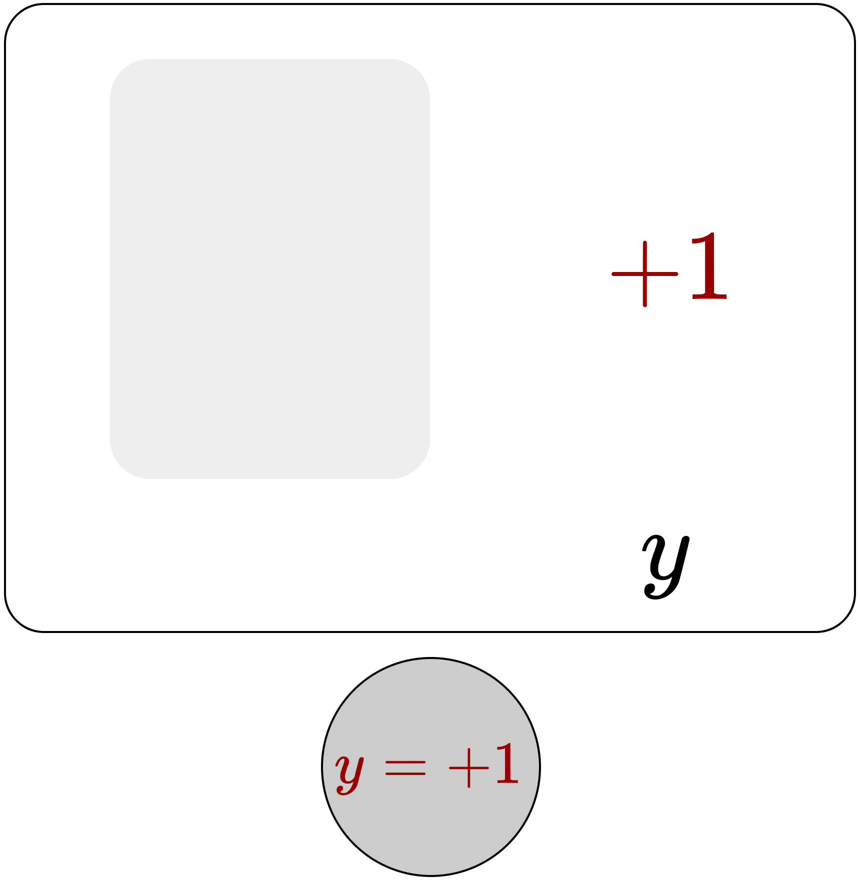
Hidden Markov Model (I)

Generative Models

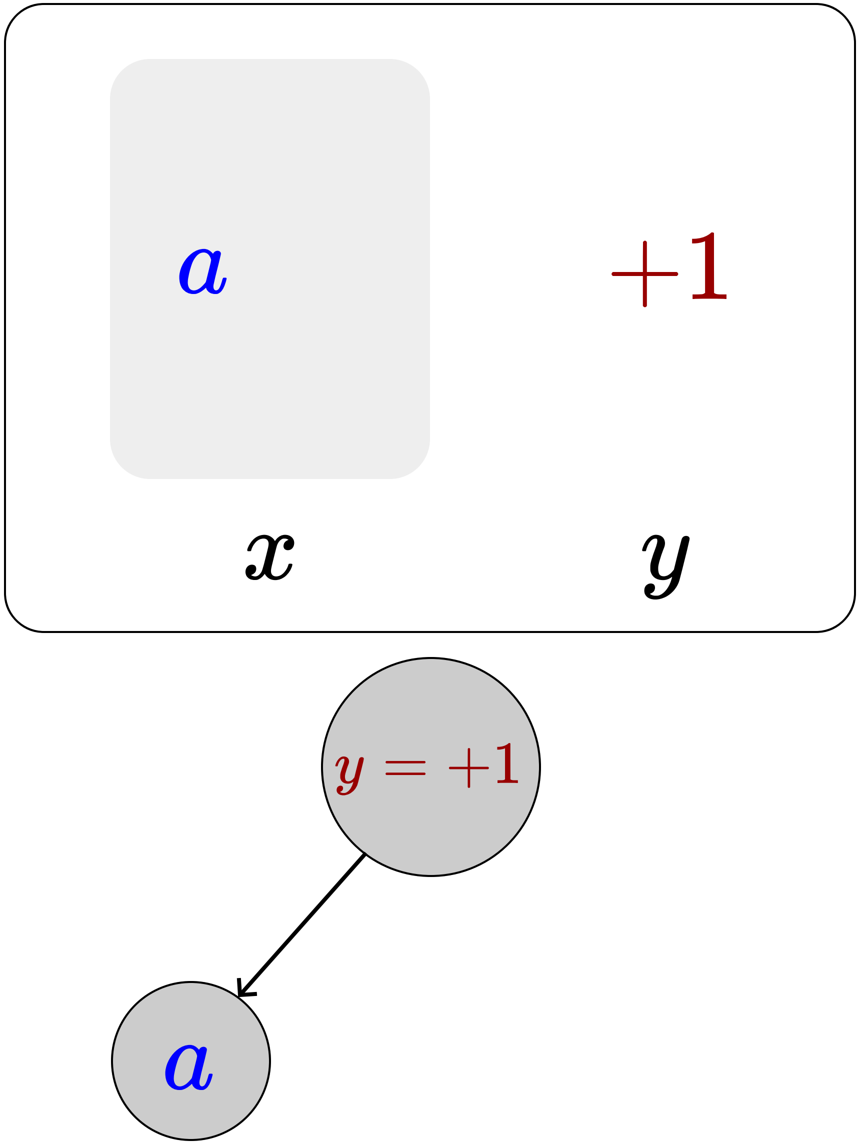
$a\ b\ c$ $\quad +1$

x $\quad y$

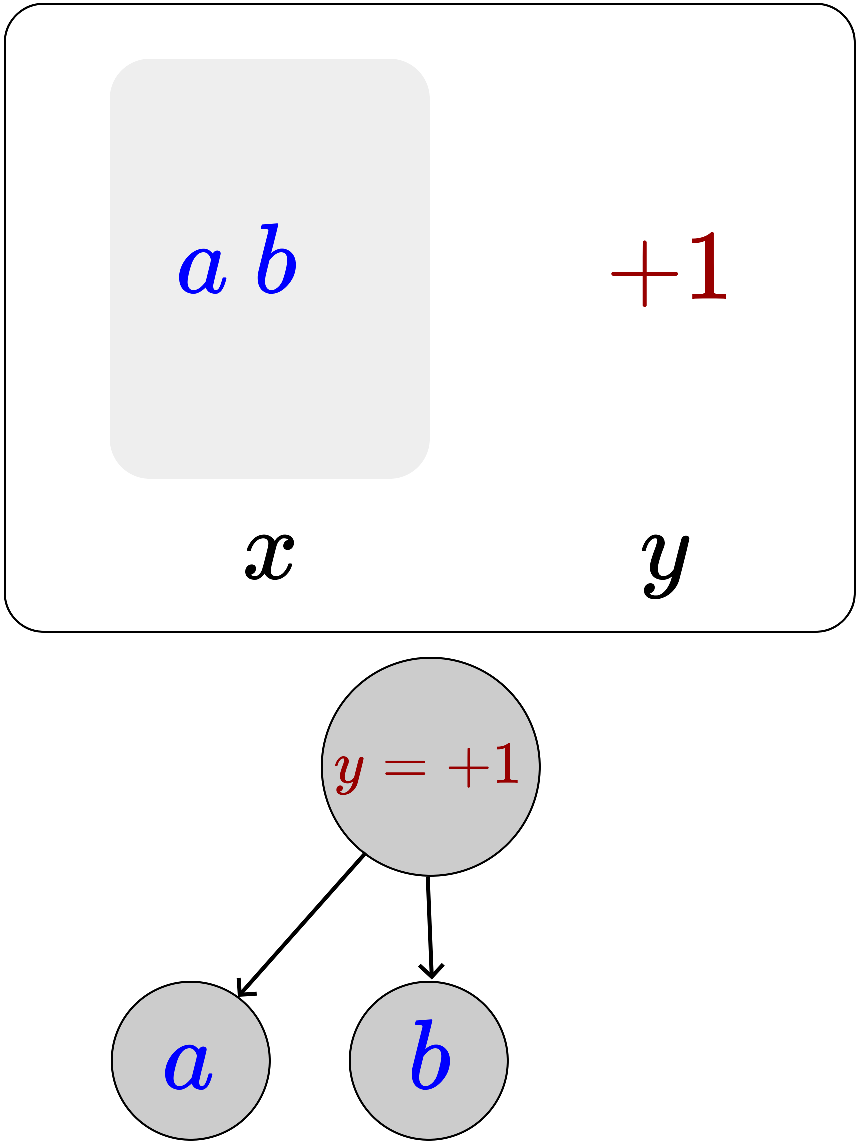
Generative Models



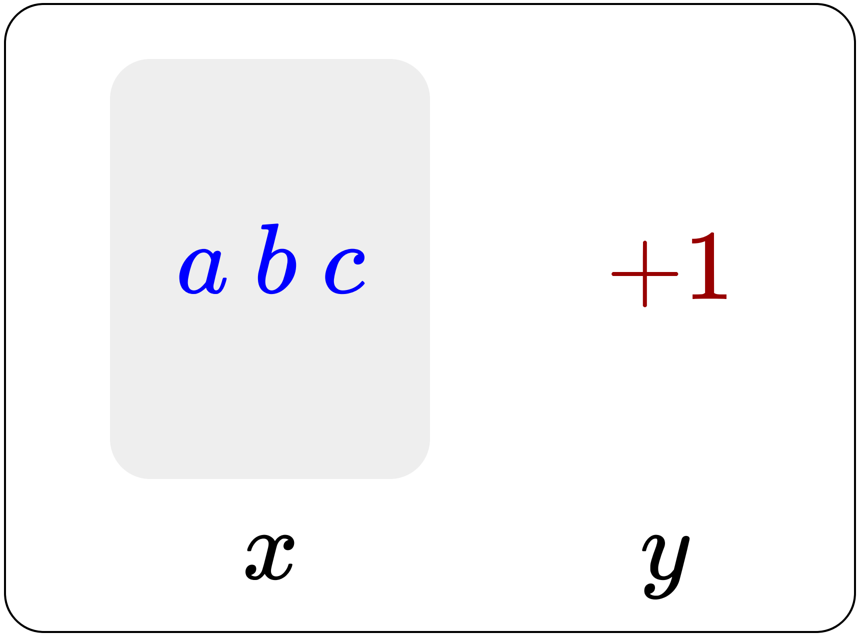
Generative Models

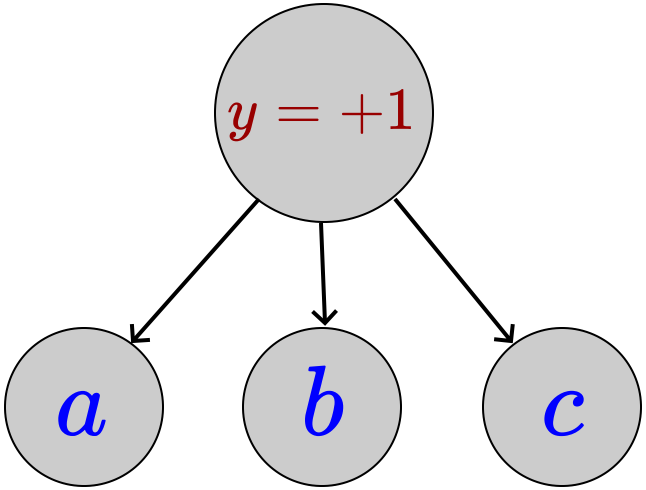


Generative Models



Generative Models





Sequence Labeling

Faith   is   a   fine   invention

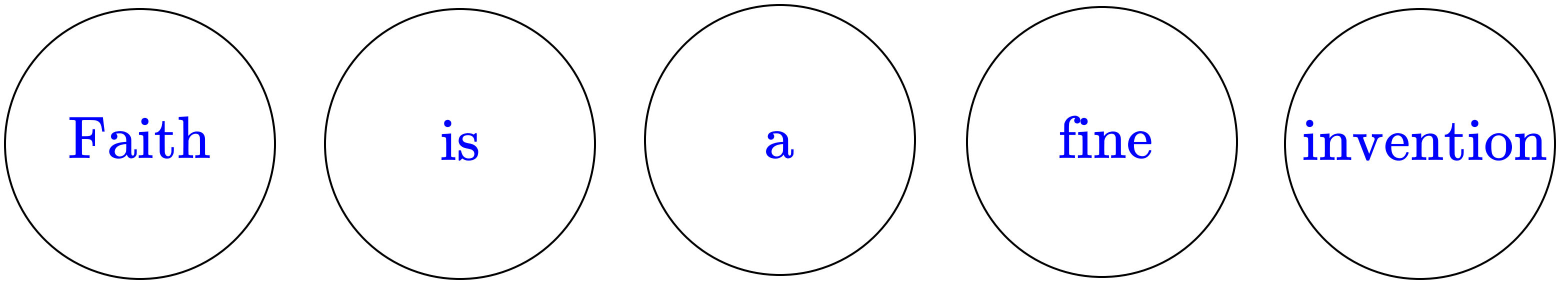
Sequence Labeling

Noun	Verb	Determiner	Adjective	Noun
$\mathbf{N}$	$\mathbf{V}$	$\mathbf{D}$	$\mathbf{A}$	$\mathbf{N}$
Faith	is	a	fine	invention

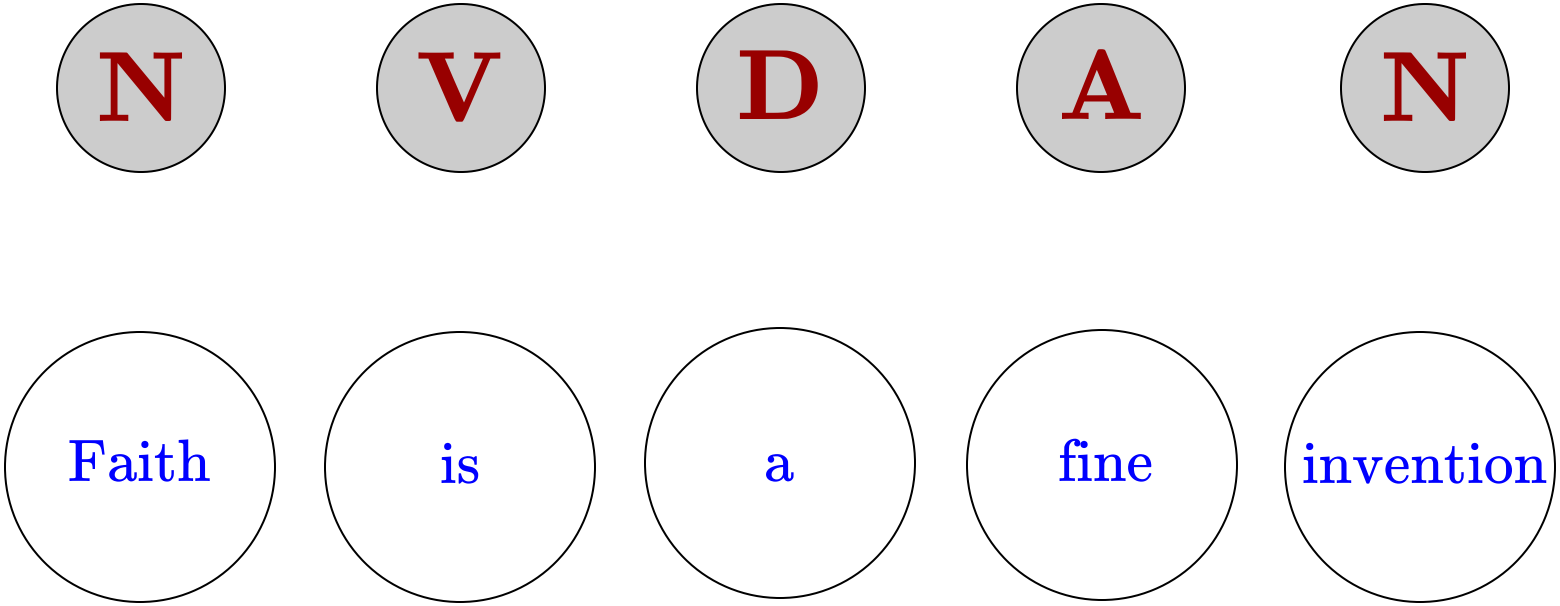
Sequence Labeling

Sequence Labeling

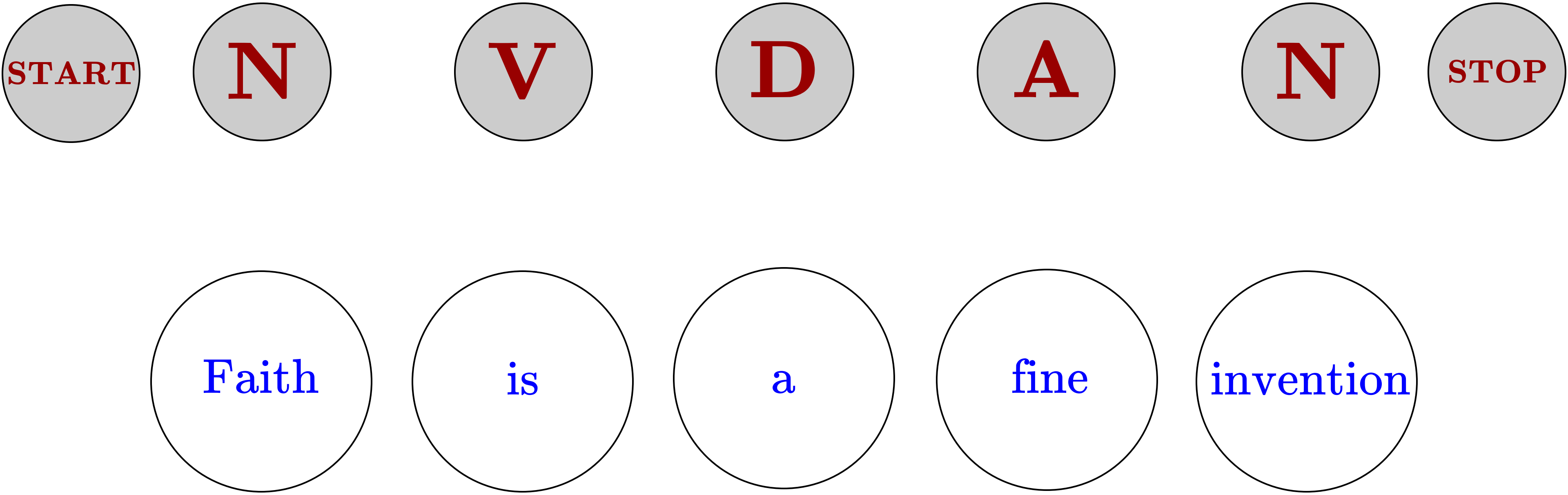
Sequence Labeling



Sequence Labeling

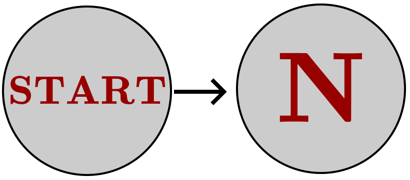


Sequence Labeling

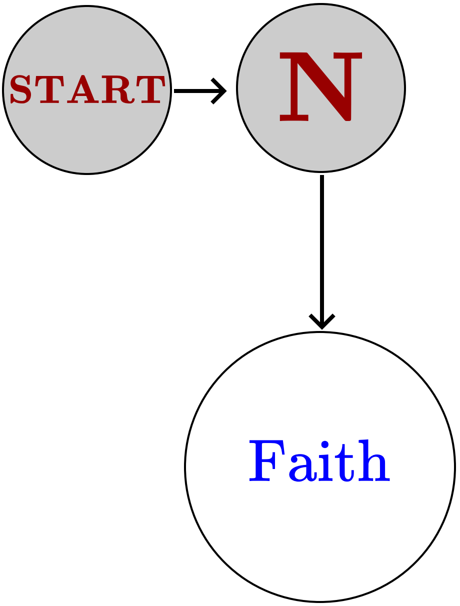


Sequence Labeling

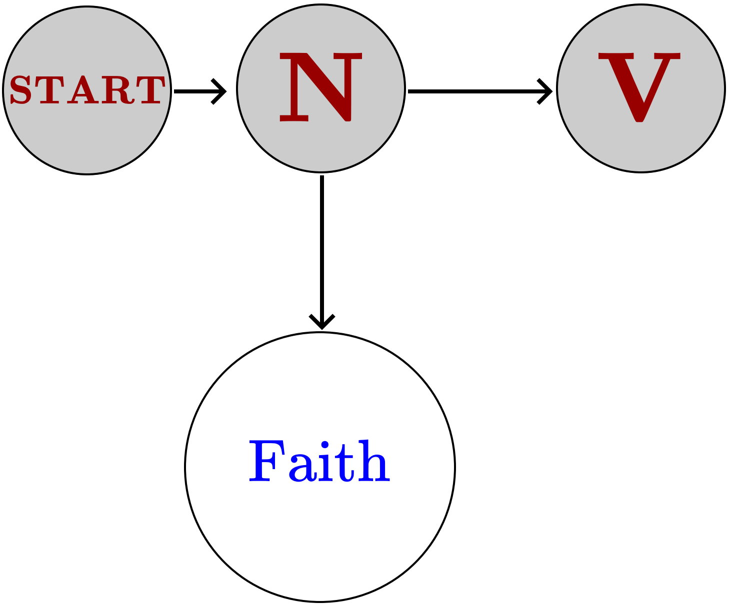
Sequence Labeling



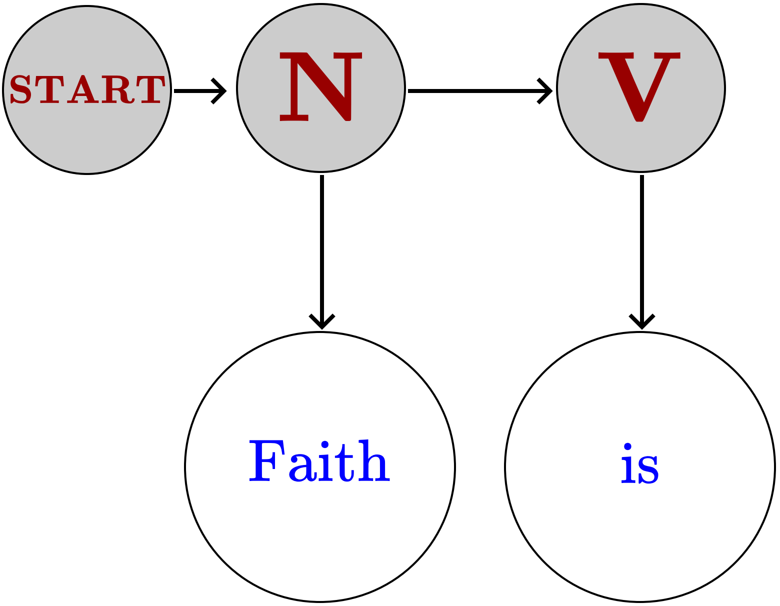
Sequence Labeling



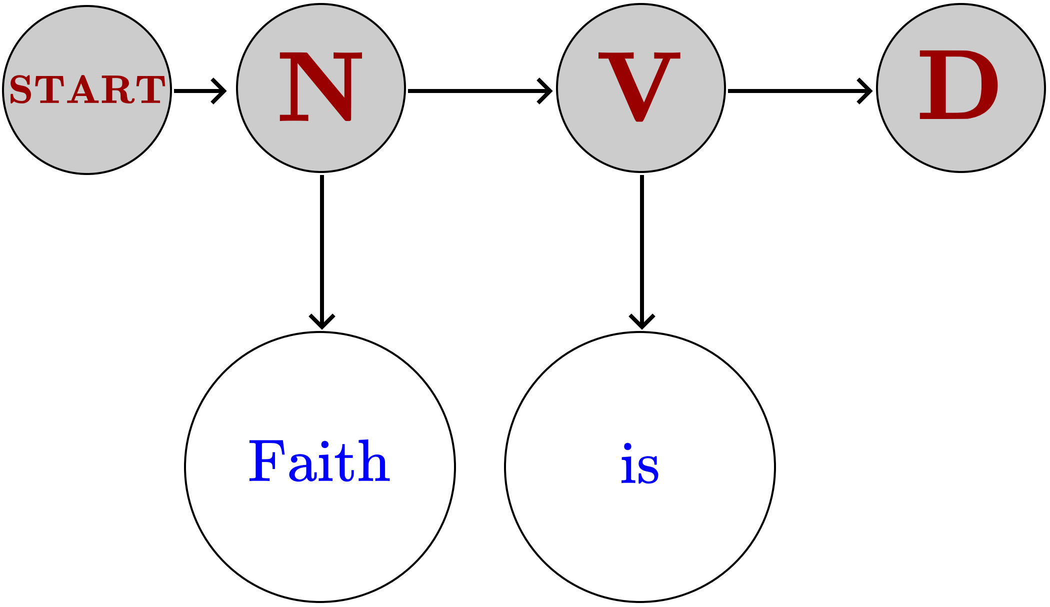
Sequence Labeling



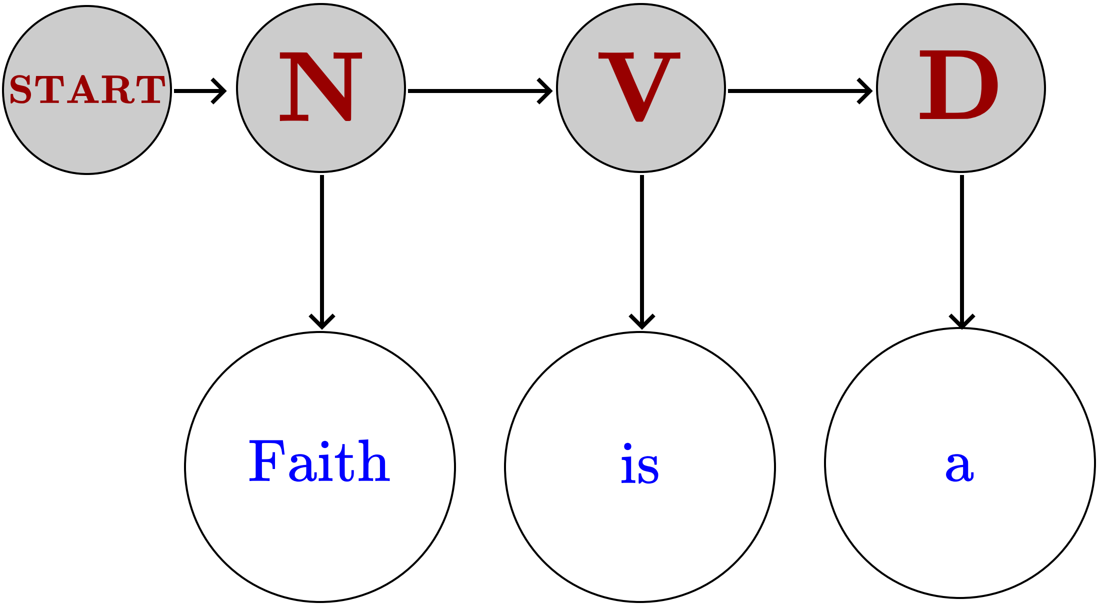
Sequence Labeling



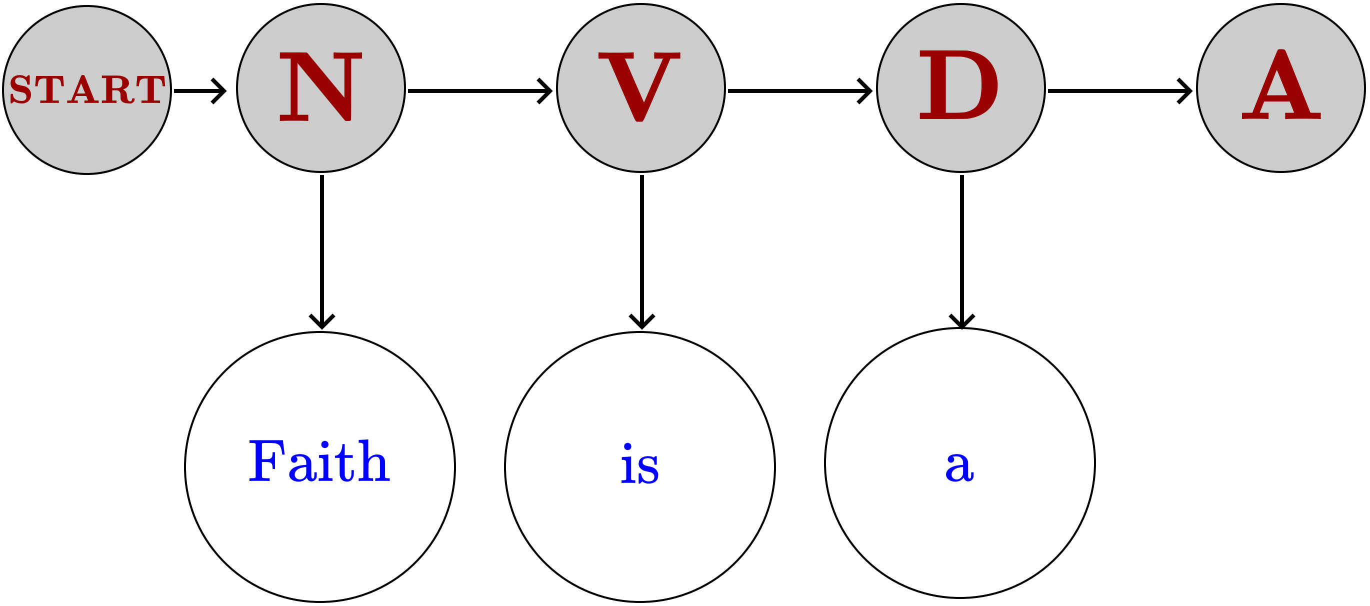
Sequence Labeling



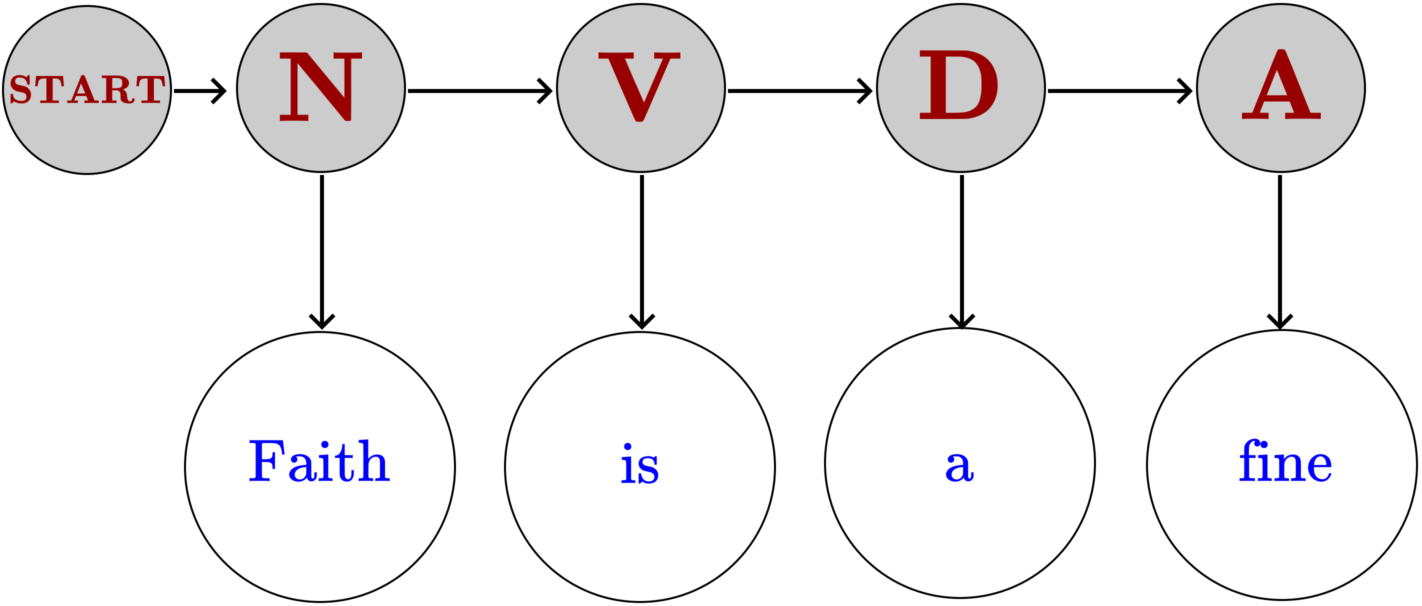
Sequence Labeling



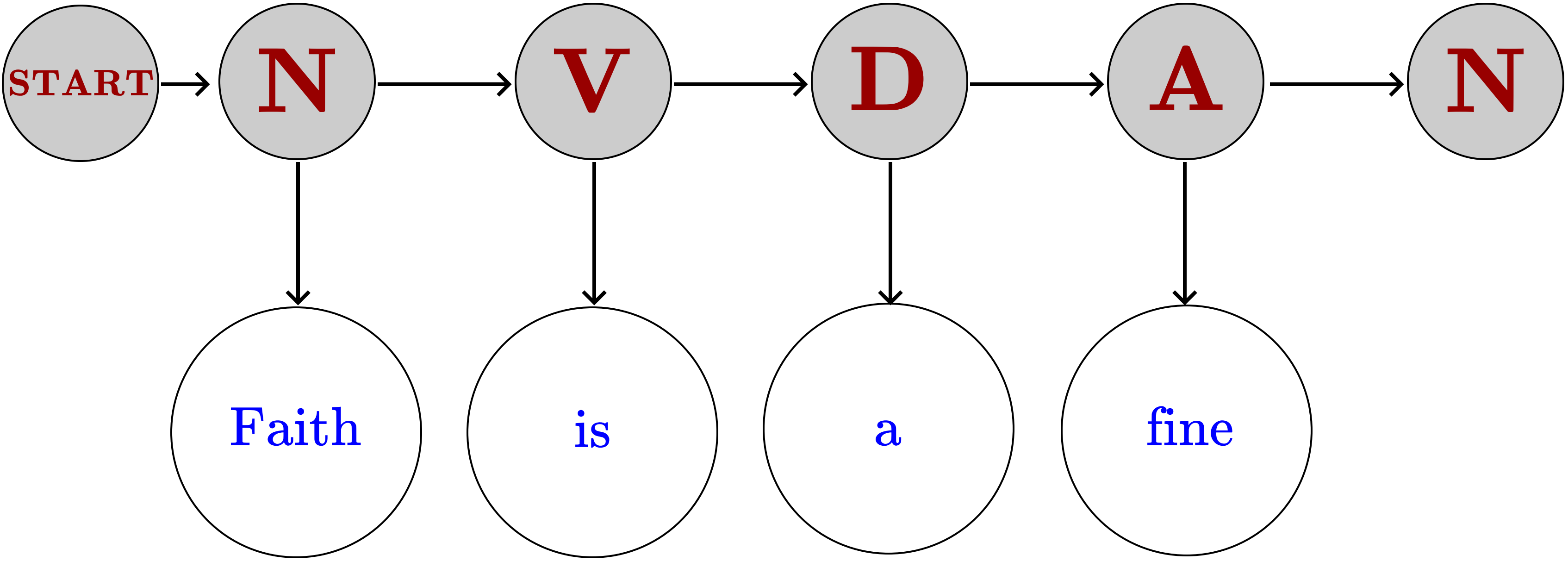
Sequence Labeling



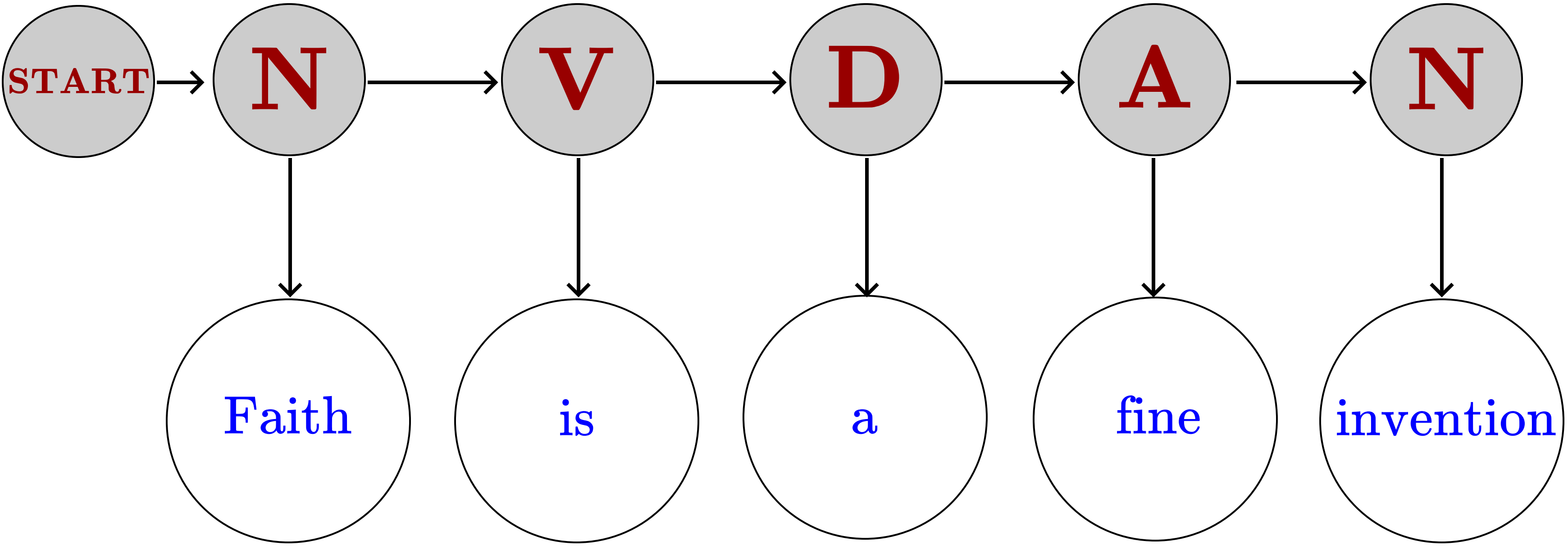
Sequence Labeling



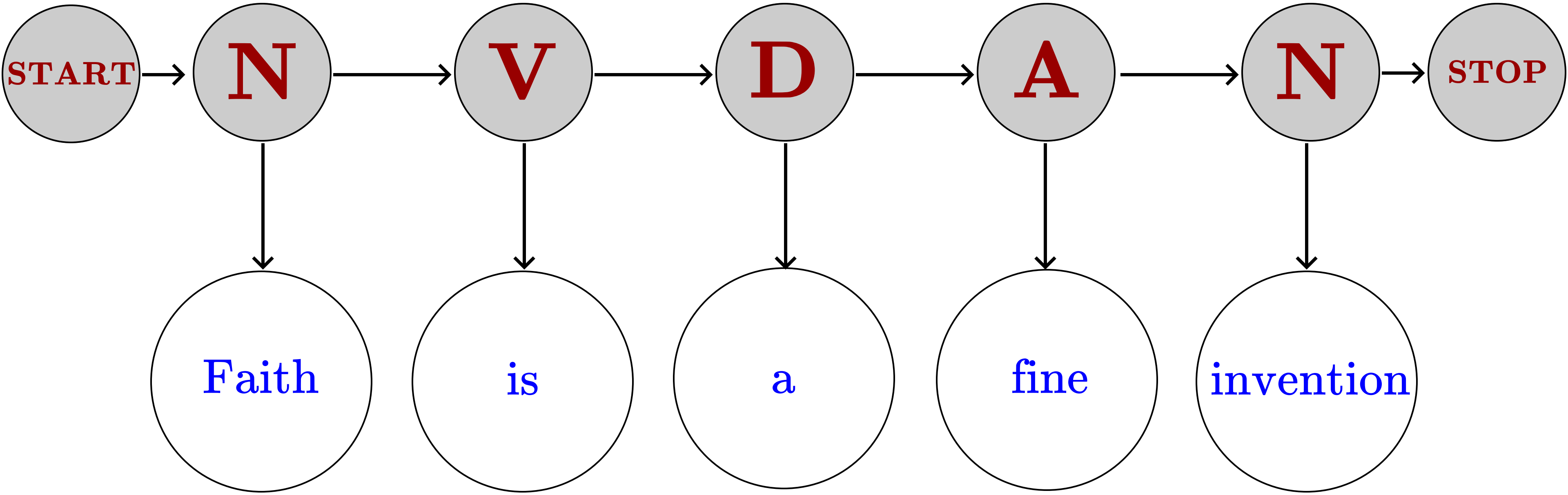
Sequence Labeling



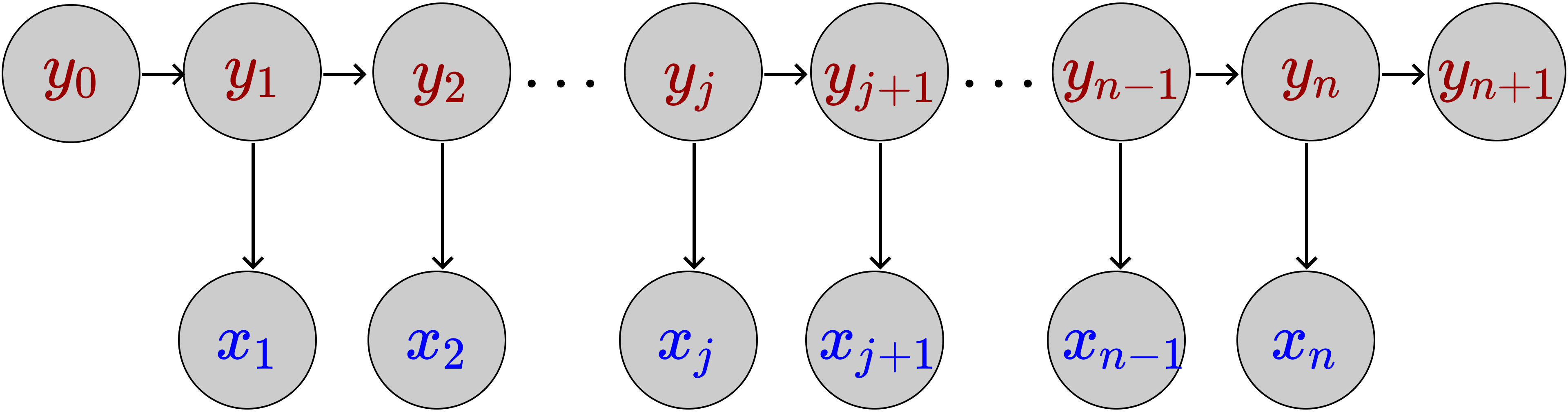
Sequence Labeling



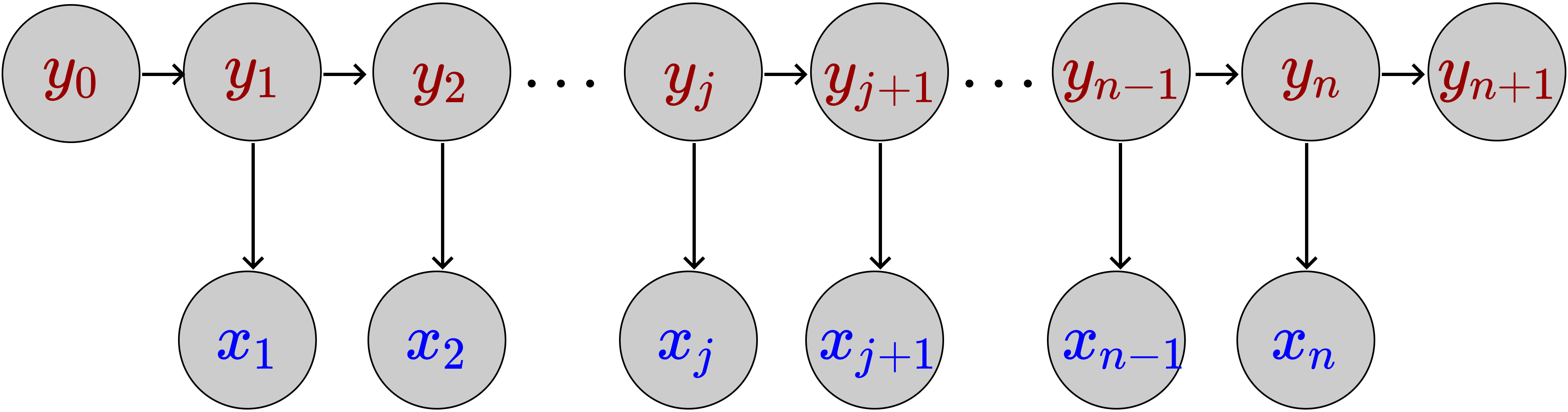
Sequence Labeling



Sequence Labeling



Sequence Labeling

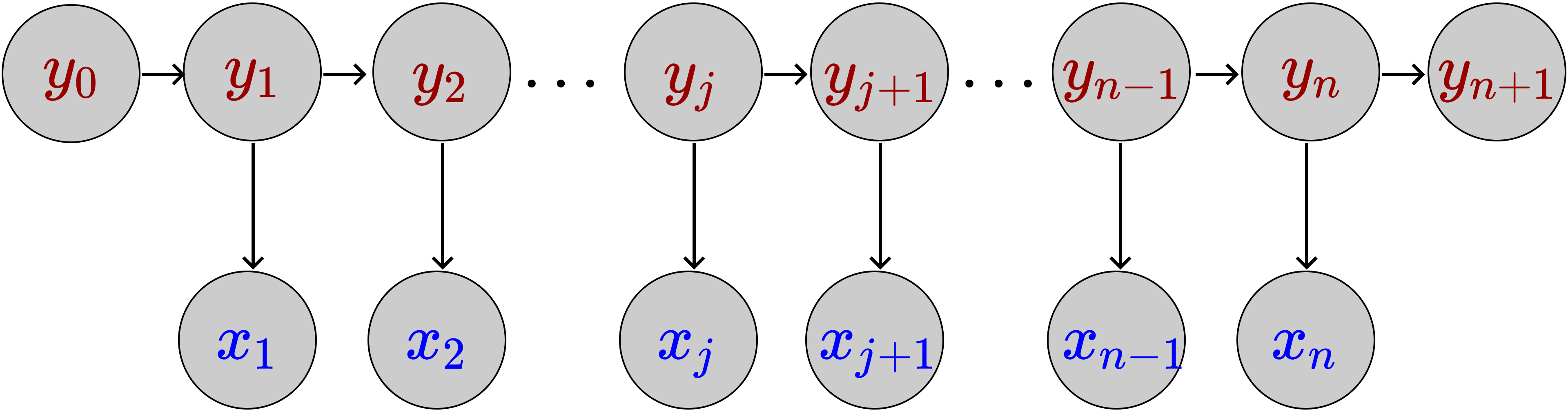


$$p(x_1, x_2, \ldots, x_{n-1}, x_n, y_0, y_1, y_2, \ldots, y_{n-1}, y_n, y_{n+1})$$

START

STOP

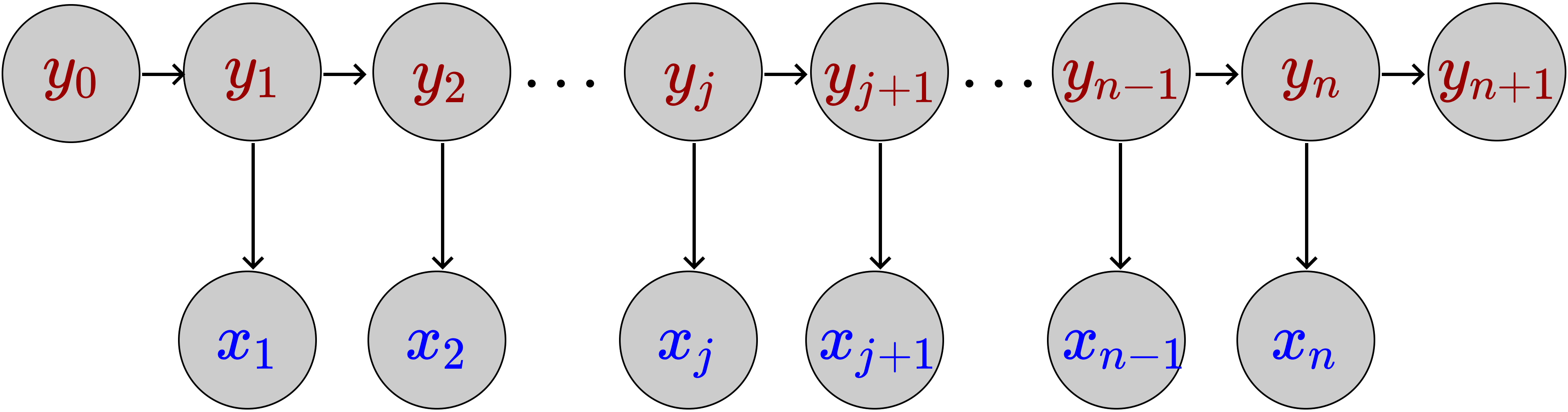
Sequence Labeling



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

$$\prod_{j=0}^{n} p(y_{j+1} | y_j) \times \prod_{j=1}^{n} p(x_j | y_j)$$

Hidden Markov Model



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

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

Hidden Markov Model

An HMM is defined by a tuple $\langle \mathcal{T}, \mathcal{O}, \theta \rangle$, where

$\mathcal{T}$

a set of states including `START` and `STOP` states.

$\mathcal{O}$

a set of observation symbols

θ

Transition and emission parameters $a_{u,v}$ and $b_u(o)$.

Hidden Markov Model
An Example

$$\mathcal{T} = \{\texttt{START}, A, B, \texttt{STOP}\}$$

$$\mathcal{O} = \{\text{``the''}, \text{``dog''}\}$$

$u \backslash v$	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

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

$$a_{u,v}$$

$$b_u(o)$$

Hidden Markov Model

An Example

$a_{u,v}$

$u \backslash v$	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

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

What is $p(\mathbf{x}, \mathbf{y})$?

Hidden Markov Model

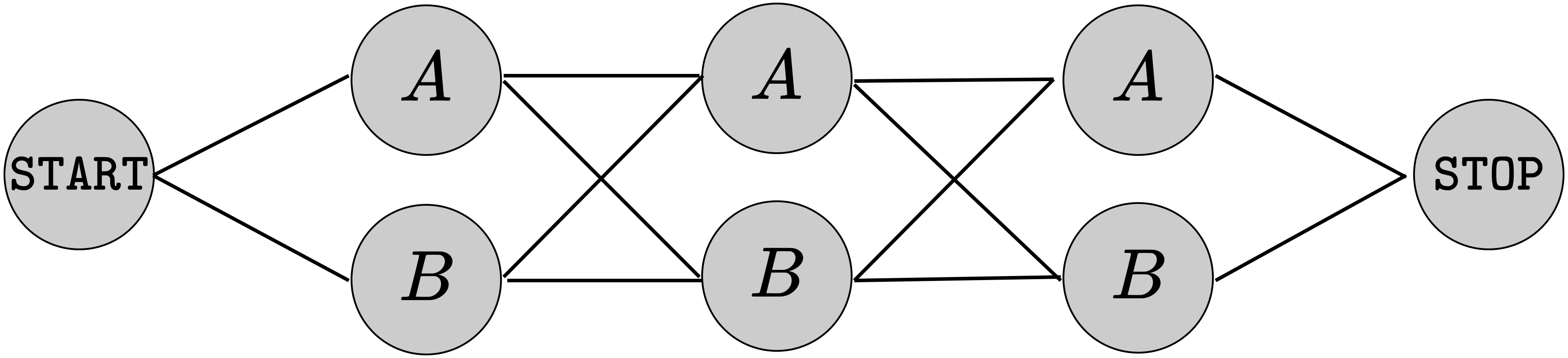
$$(\mathbf{x}, \mathbf{y}) = \text{the}/A, \text{dog}/B, \text{the}/A$$

$a_{u,v}$

$u \backslash v$	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



One path corresponds to one label sequence.

Hidden Markov Model

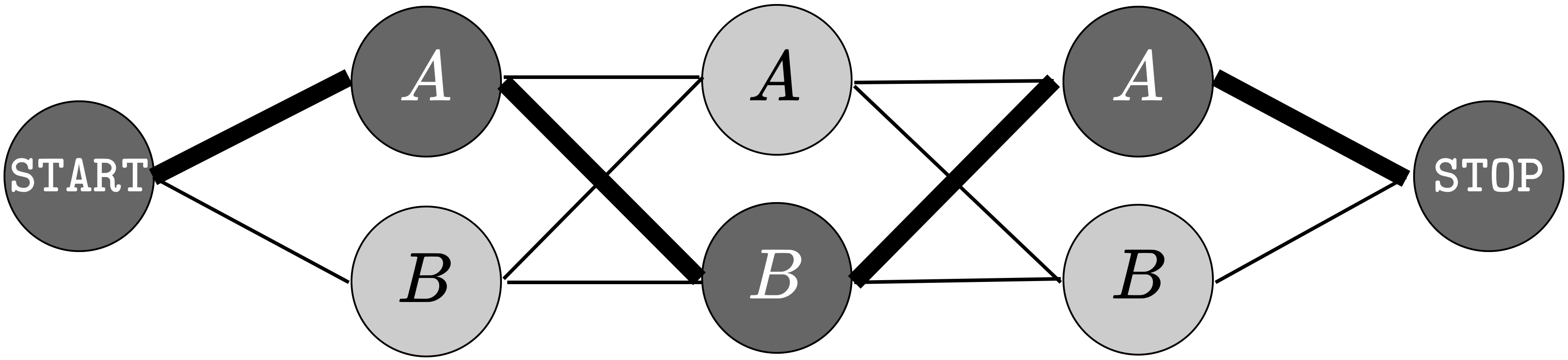
$$(\mathbf{x}, \mathbf{y}) = \text{the}/A, \text{dog}/B, \text{the}/A$$

$a_{u,v}$

$u \backslash v$	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



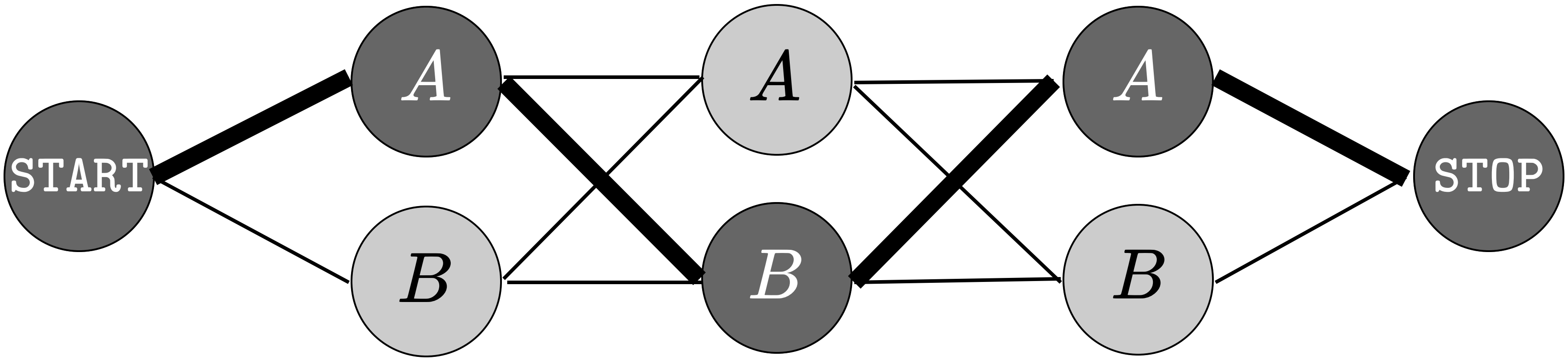
$a_{\texttt{START},A}$

Hidden Markov Model

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

$u \backslash v$	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

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



$$a_{\texttt{START},A} \times b_A(\text{``The''})$$

Hidden Markov Model

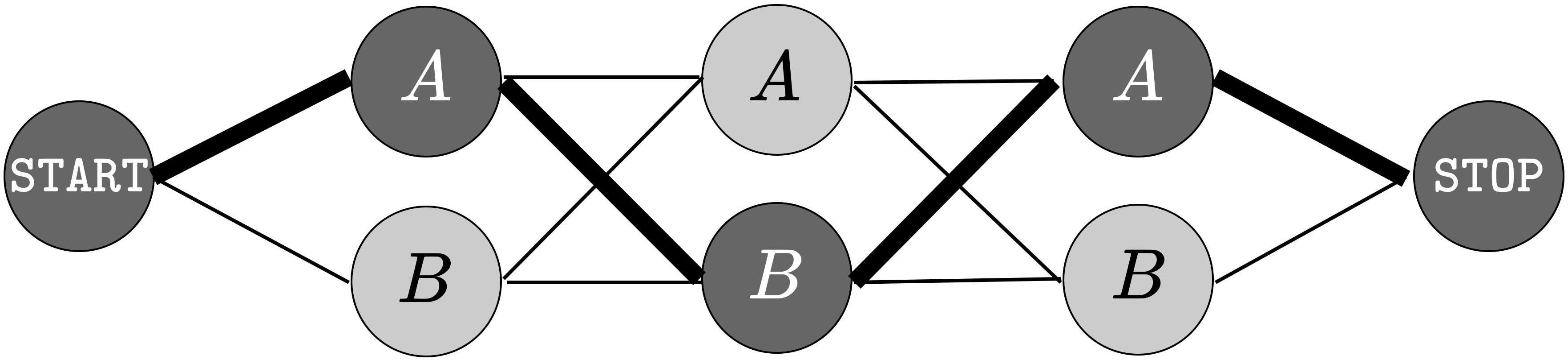
$$(\mathbf{x}, \mathbf{y}) = \text{the}/A, \text{dog}/B, \text{the}/A$$

$a_{u,v}$

$u \backslash v$	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



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

Hidden Markov Model

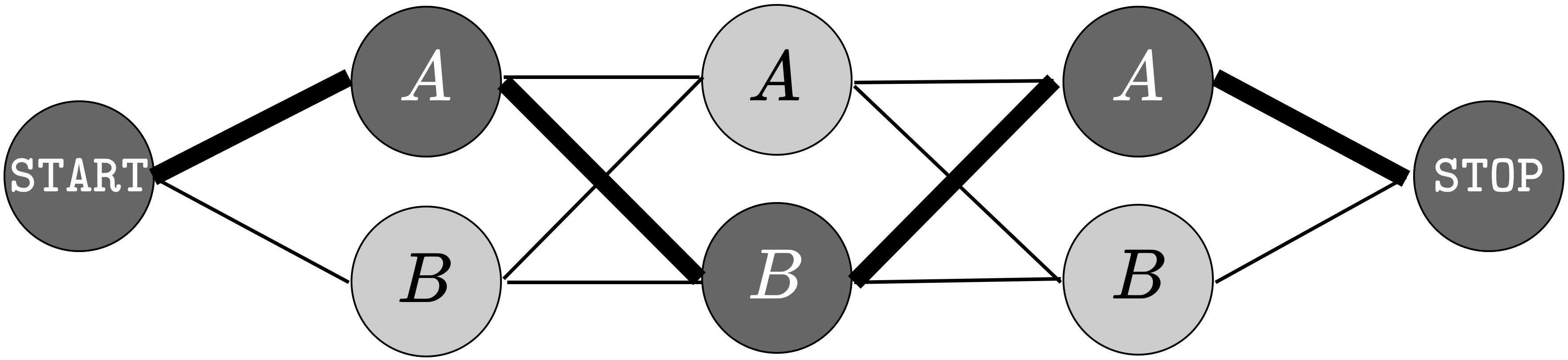
$$(\mathbf{x}, \mathbf{y}) = \text{the}/A, \text{dog}/B, \text{the}/A$$

$a_{u,v}$

$u \backslash v$	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



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

Hidden Markov Model

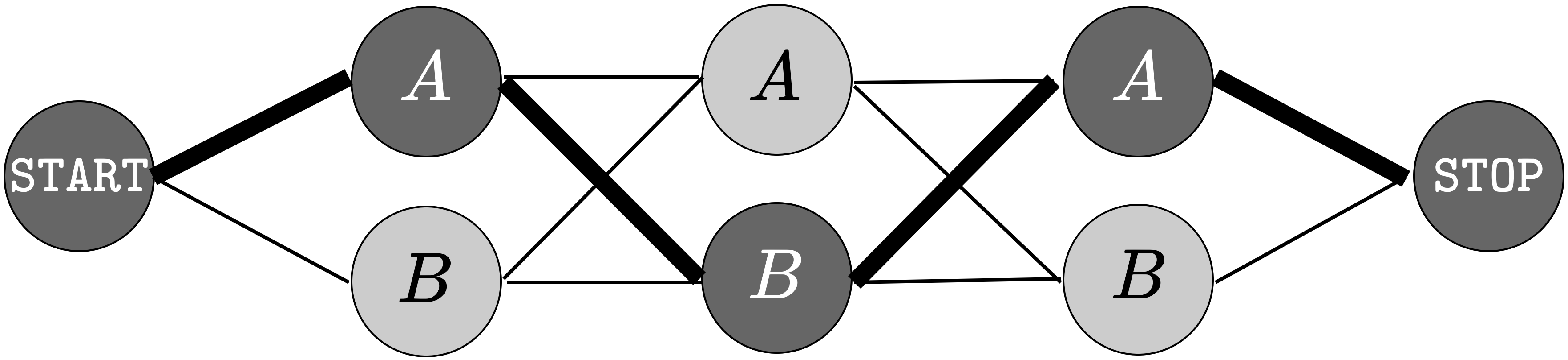
$$(\mathbf{x}, \mathbf{y}) = \text{the}/A, \text{dog}/B, \text{the}/A$$

$a_{u,v}$

$u \backslash v$	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



$$a_{\texttt{START},A} \times b_A(\text{"The"}) \times a_{A,B} \times b_B(\text{"dog"}) \times a_{B,A}$$

Hidden Markov Model

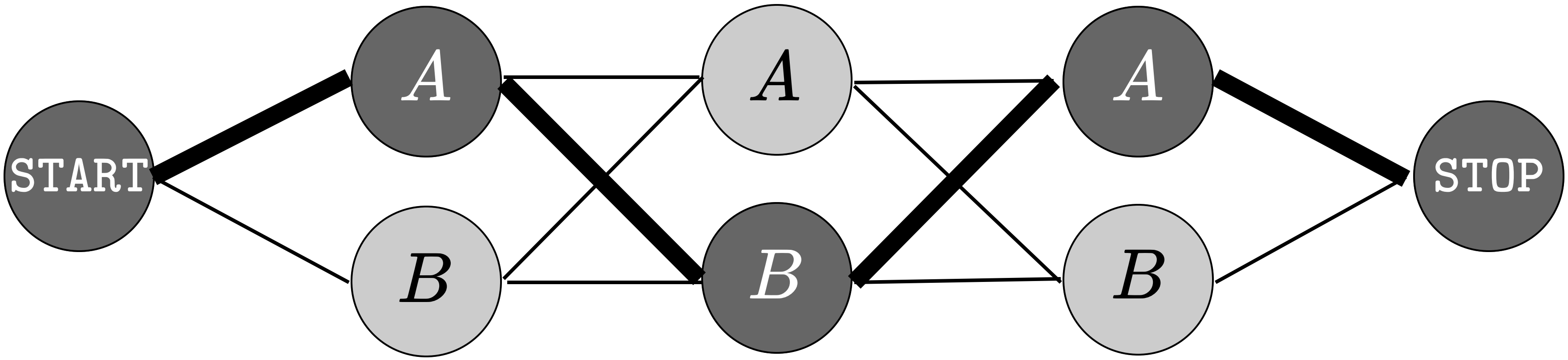
$$(\mathbf{x}, \mathbf{y}) = \text{the}/A, \text{dog}/B, \text{the}/A$$

$a_{u,v}$

$u \backslash v$	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



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

Hidden Markov Model

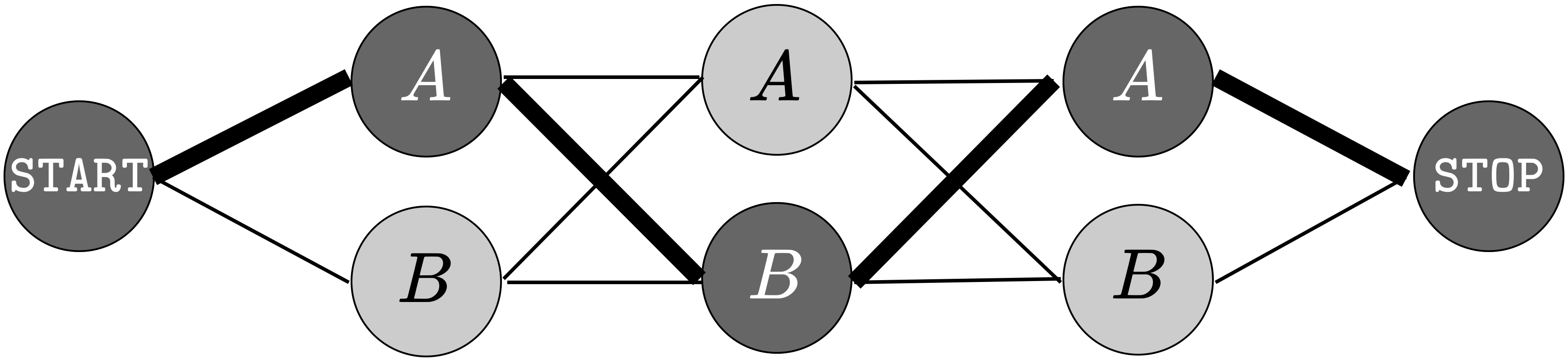
$$(\mathbf{x}, \mathbf{y}) = \text{the}/A, \text{dog}/B, \text{the}/A$$

$a_{u,v}$

$u \backslash v$	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

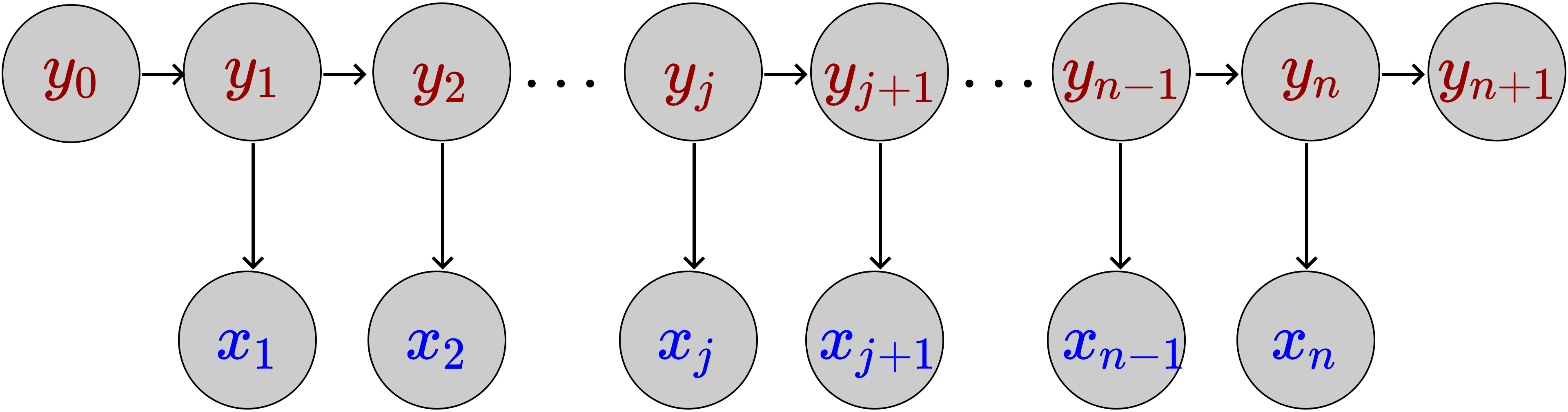
$b_u(o)$

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



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

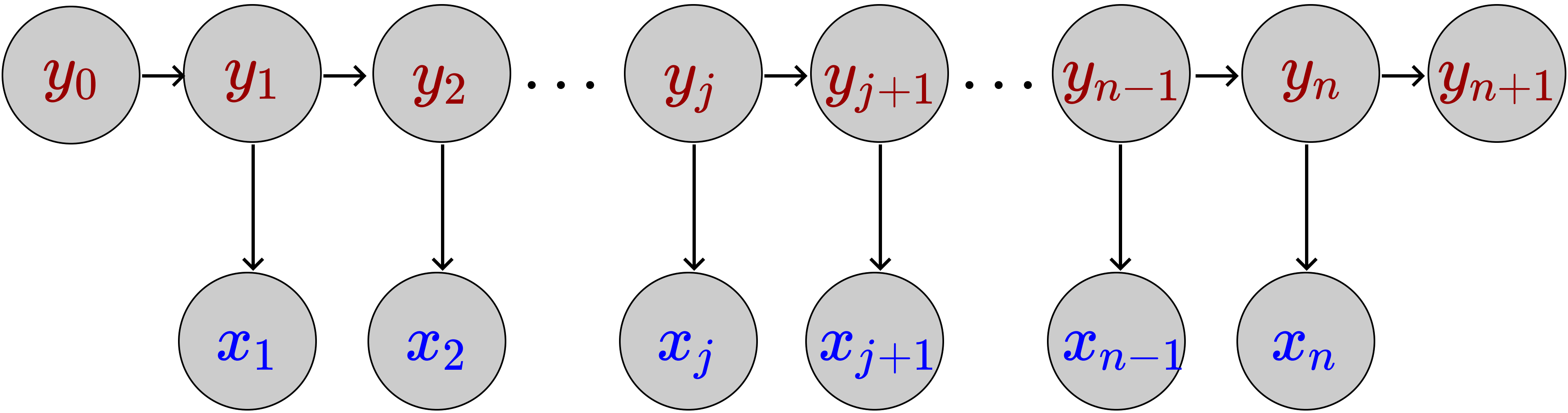
Hidden Markov Model



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

Now that we know what are the model parameters, how do we estimate them? In other words, how to do learning?

Hidden Markov Model

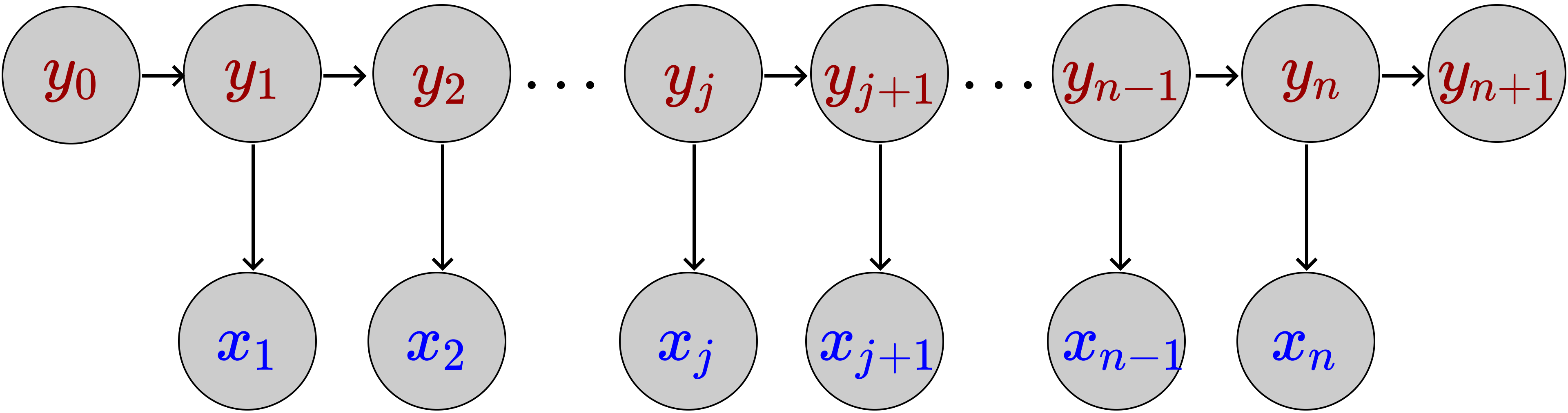


Number of times we see a transition from u to v

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

Number of times we see the state u in the training set

Hidden Markov Model



Number of times we see observation o generated from u

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

Number of times we see the state u in the training set

Hidden Markov Model

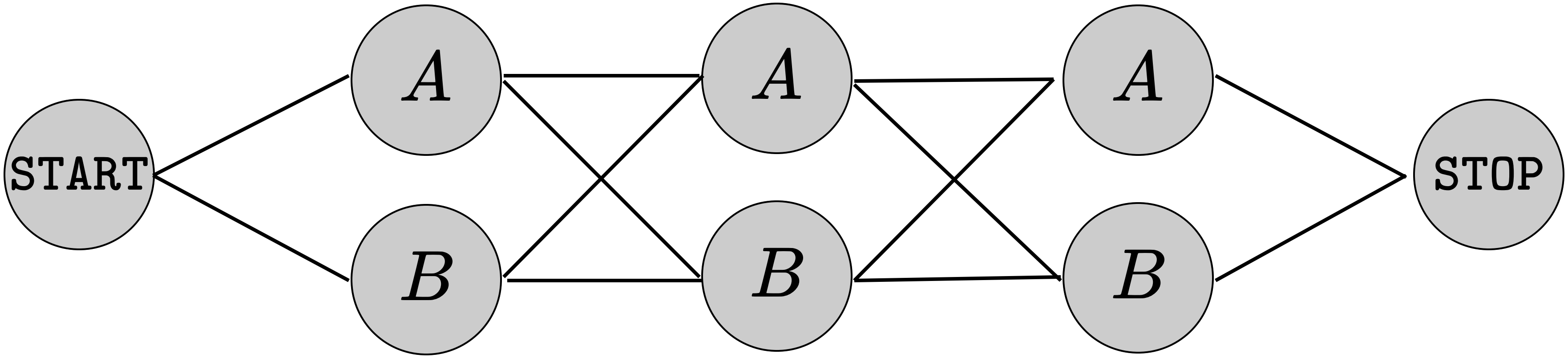
$a_{u,v}$

$u \backslash v$	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

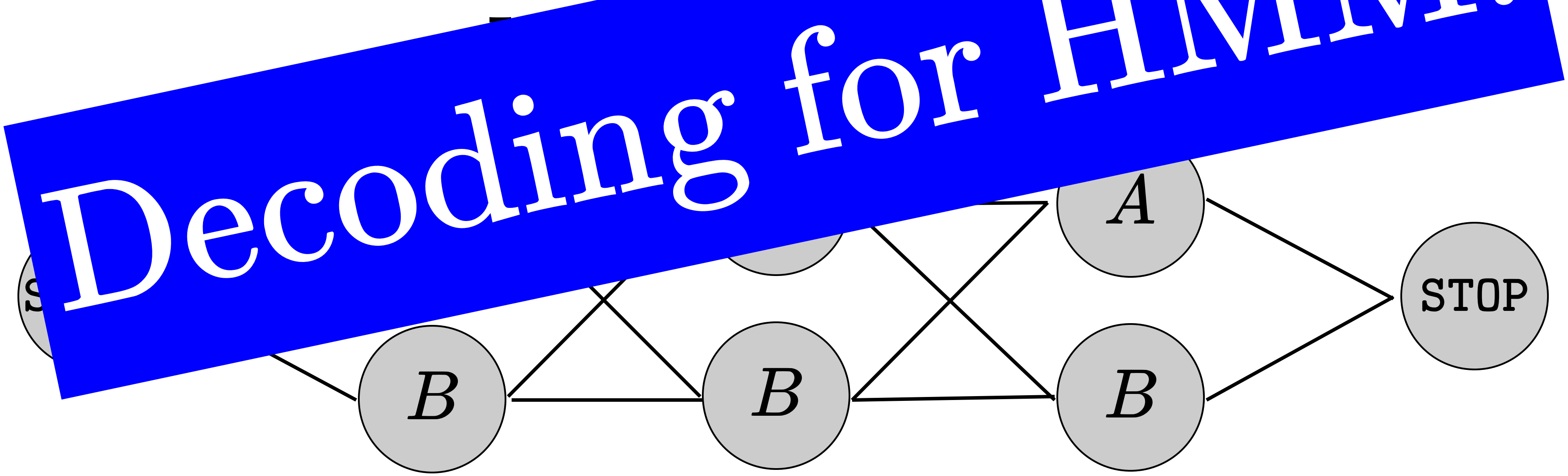
$a_{u,v}$

$u \backslash v$	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	[illegible]

Decoding for HMM!



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