50.007 Machine Learning

Hidden Markov Model

(Adapted from Prof. Lu Wei's slides)

Roy Ka-Wei Lee

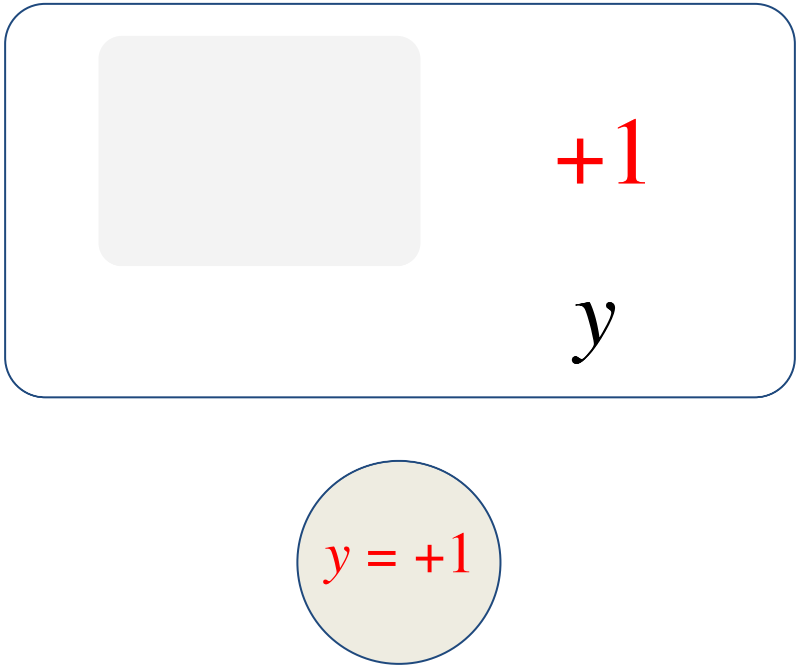
Assistant Professor, DAI/ISTD, SUTD

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

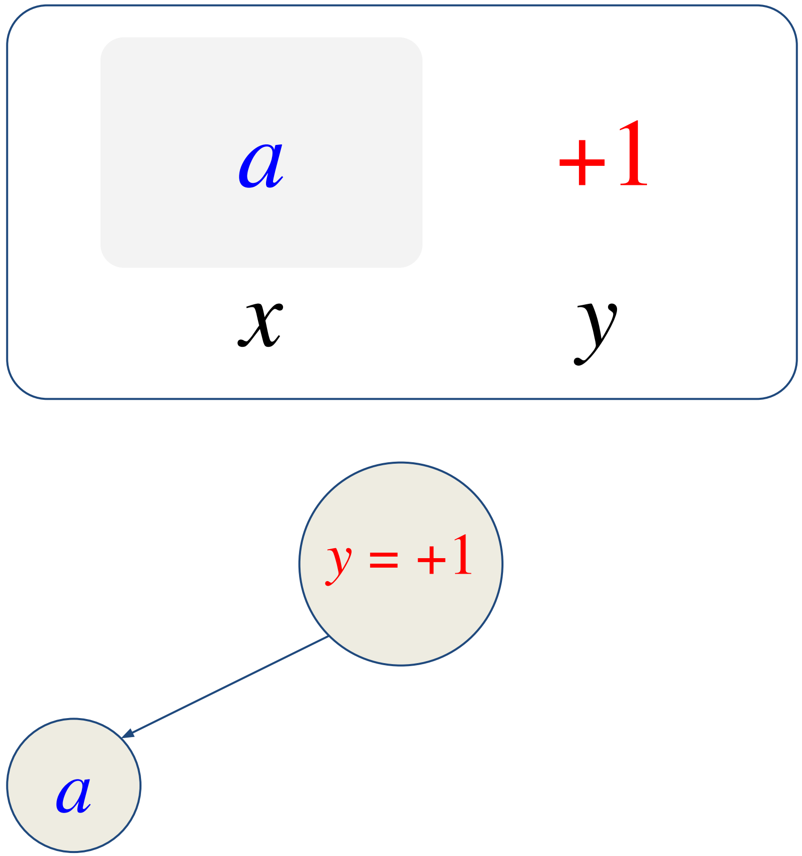
Generative Models

x	y
$a\ b\ c$	$+1$

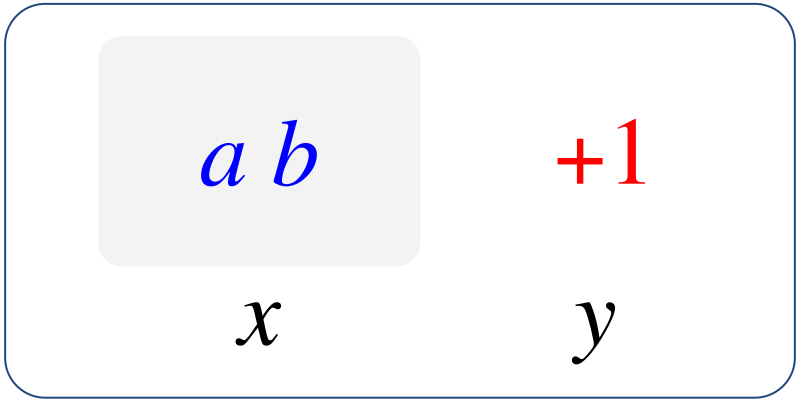
Generative Models

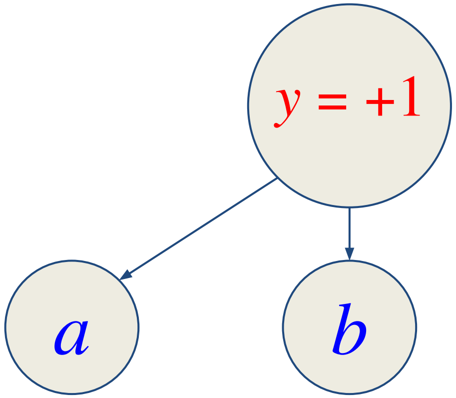


Generative Models

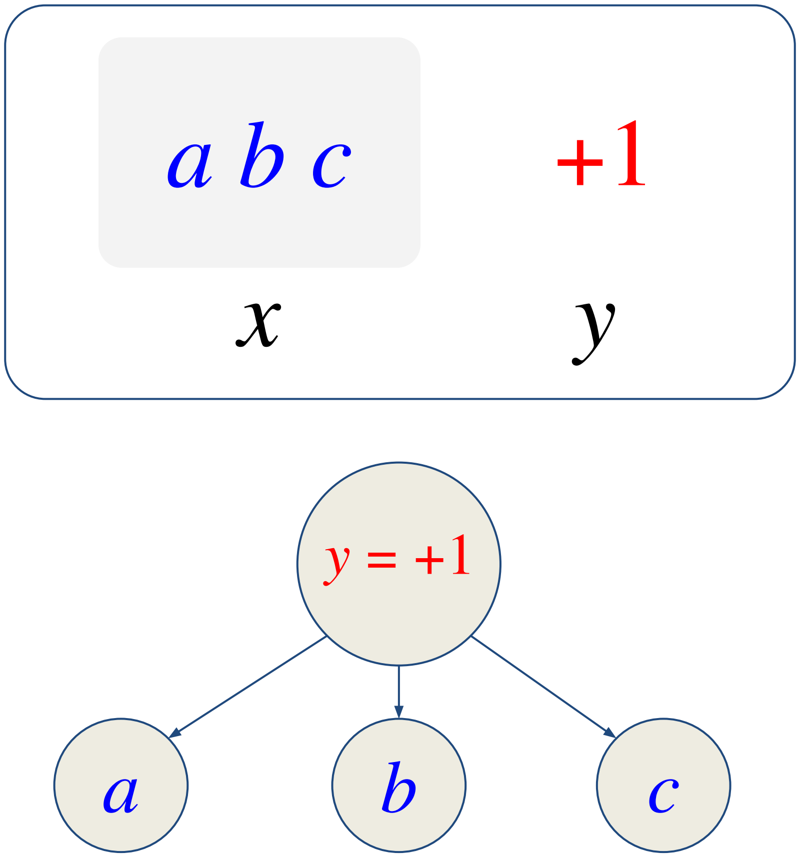


Generative Models

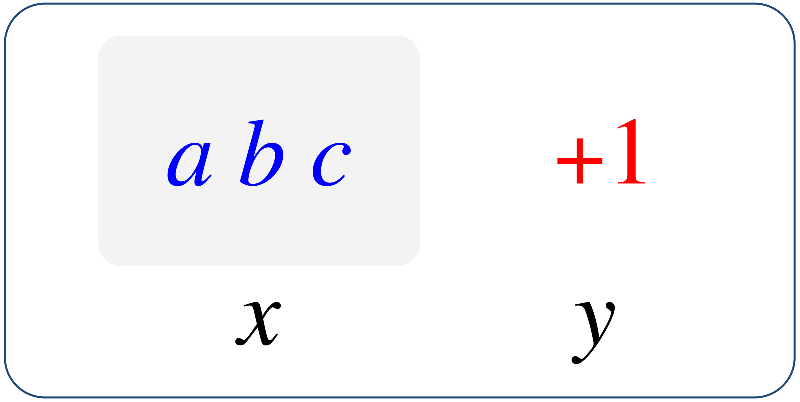


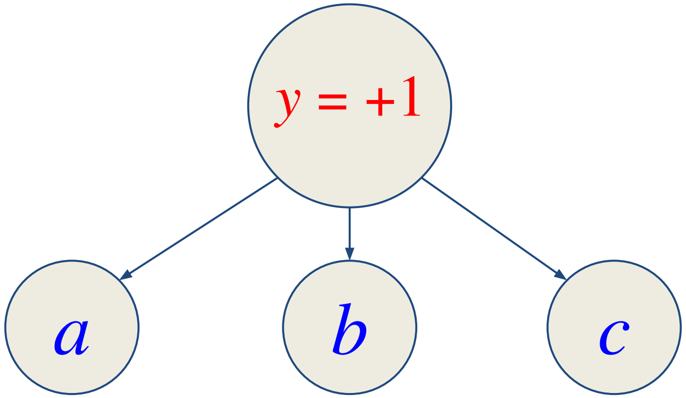


Generative Models



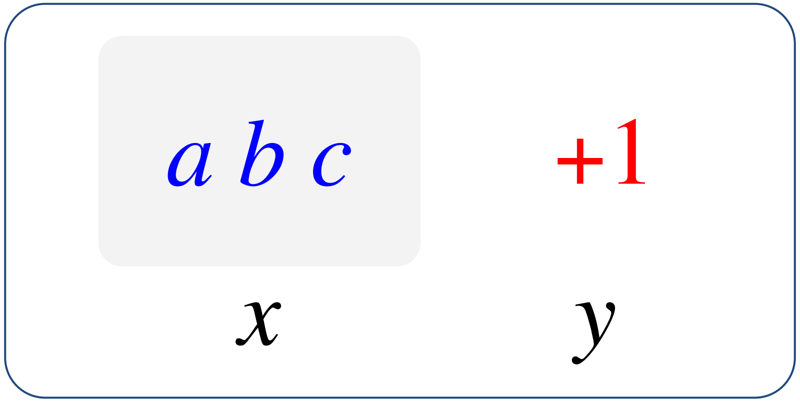
Generative Models

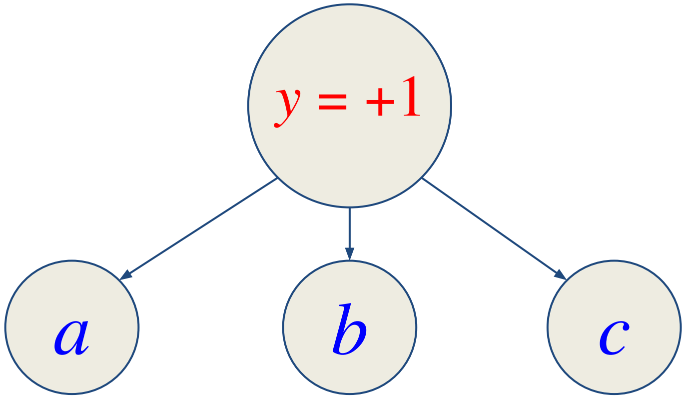




$p(x\text{=“}a, b, c\text{”}, y = +1)$

Generative Models





$$p(x=\text{“}a, b, c\text{”}, y = +1) = p(y = +1)\, p(a \mid y = +1)\, p(b \mid y = +1)\, p(c \mid y = +1)$$

Sequence Labeling

Faith is a fine invention

Sequence Labeling

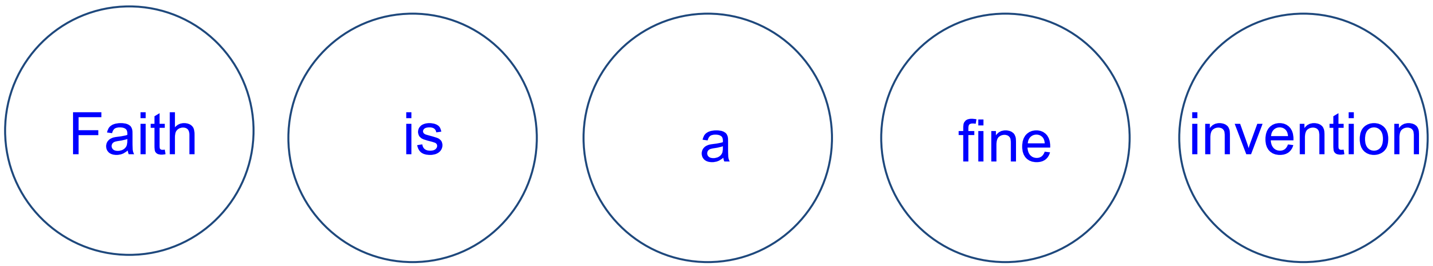
Noun	Verb	Determiner	Adjective	Noun
N	**V**	**D**	**A**	**N**
Faith	is	a	fine	invention

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

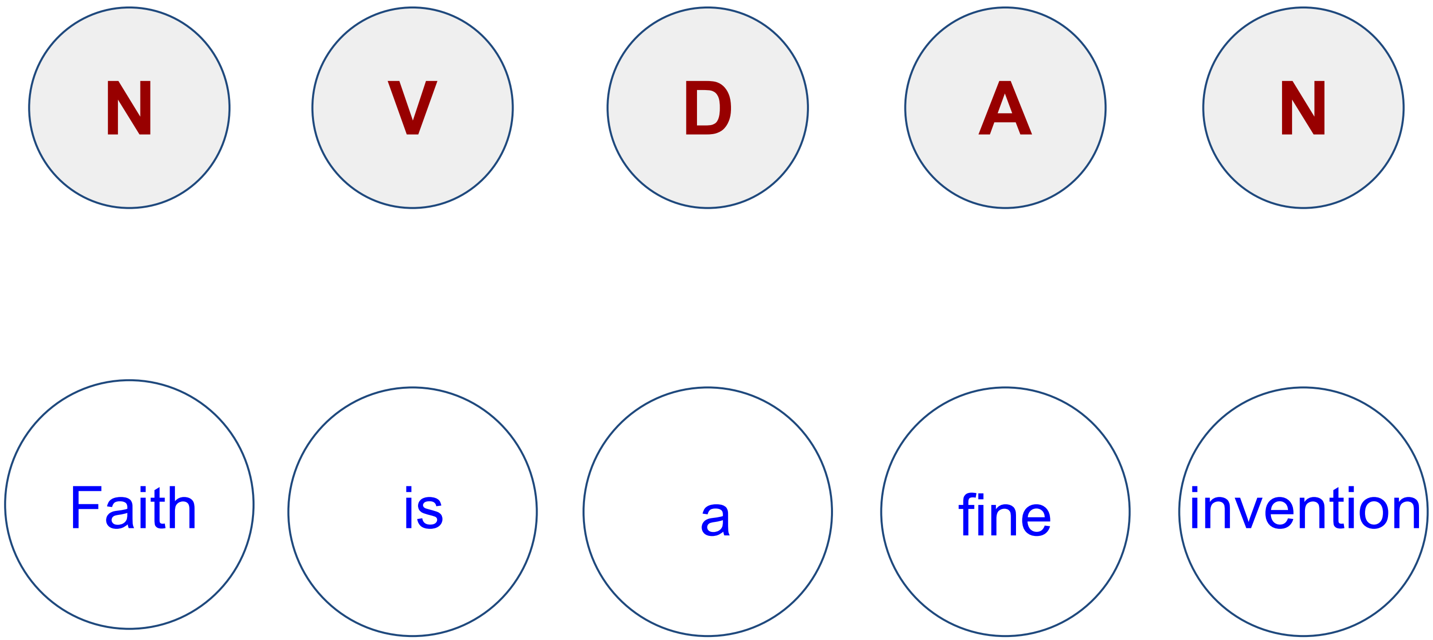
Sequence Labeling

Sequence Labeling

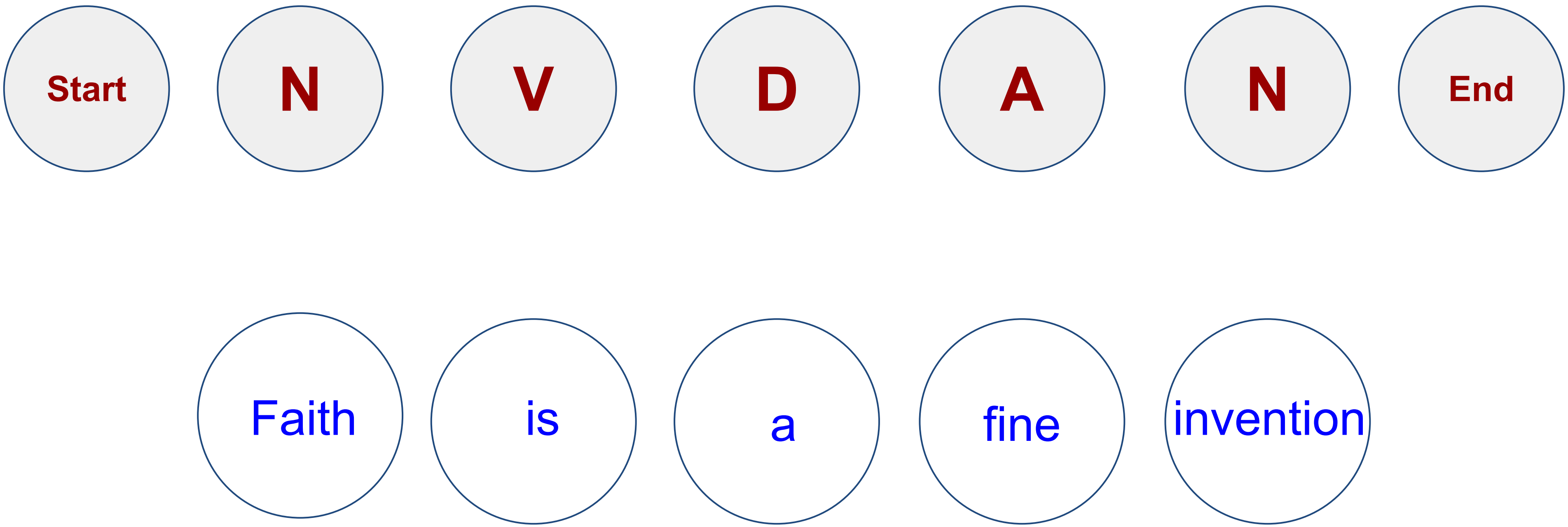
Sequence Labeling



Sequence Labeling



Sequence Labeling

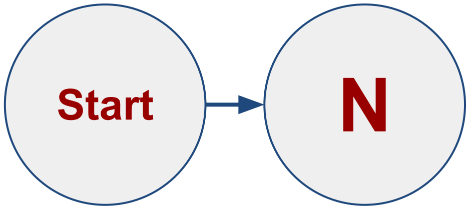


Sequence Labeling

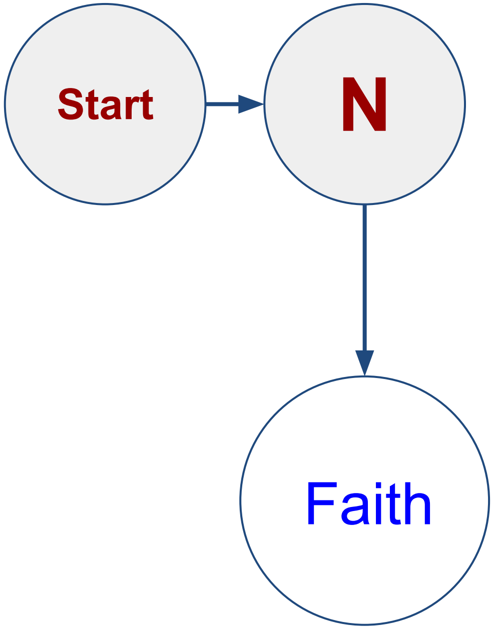
Start

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

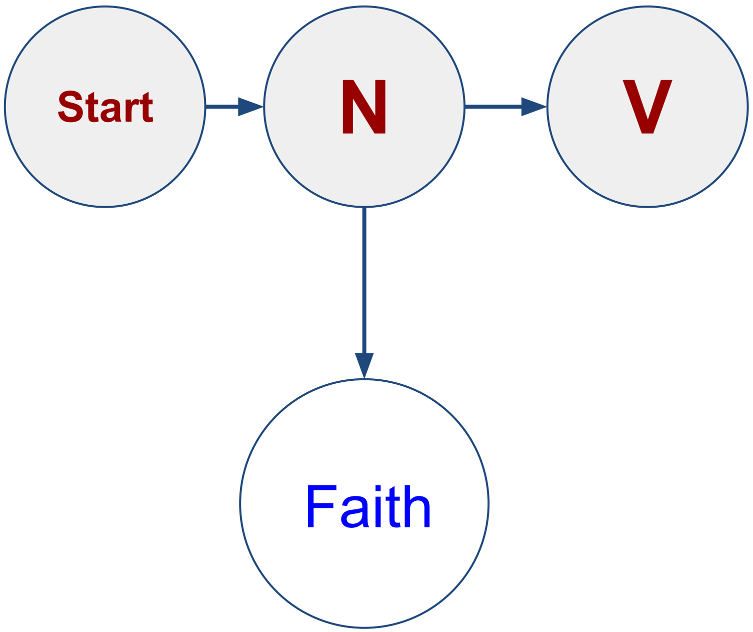
Sequence Labeling



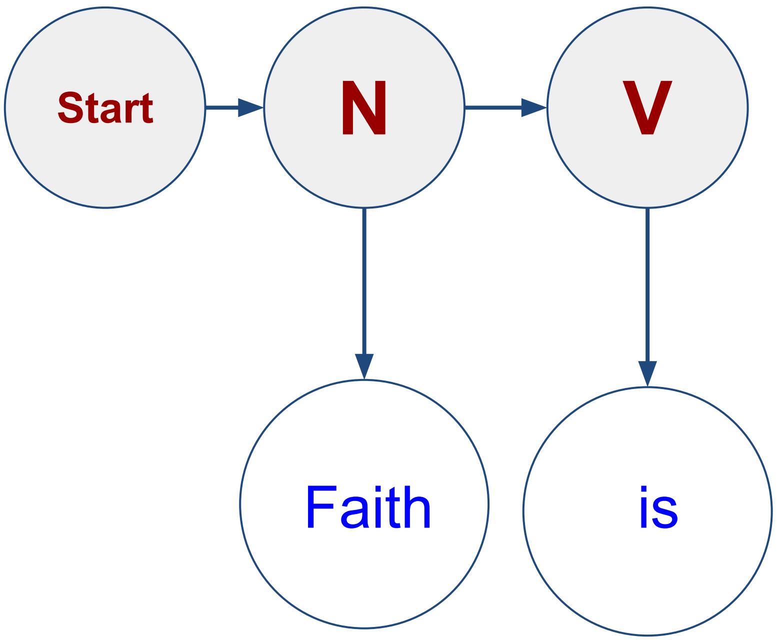
Sequence Labeling



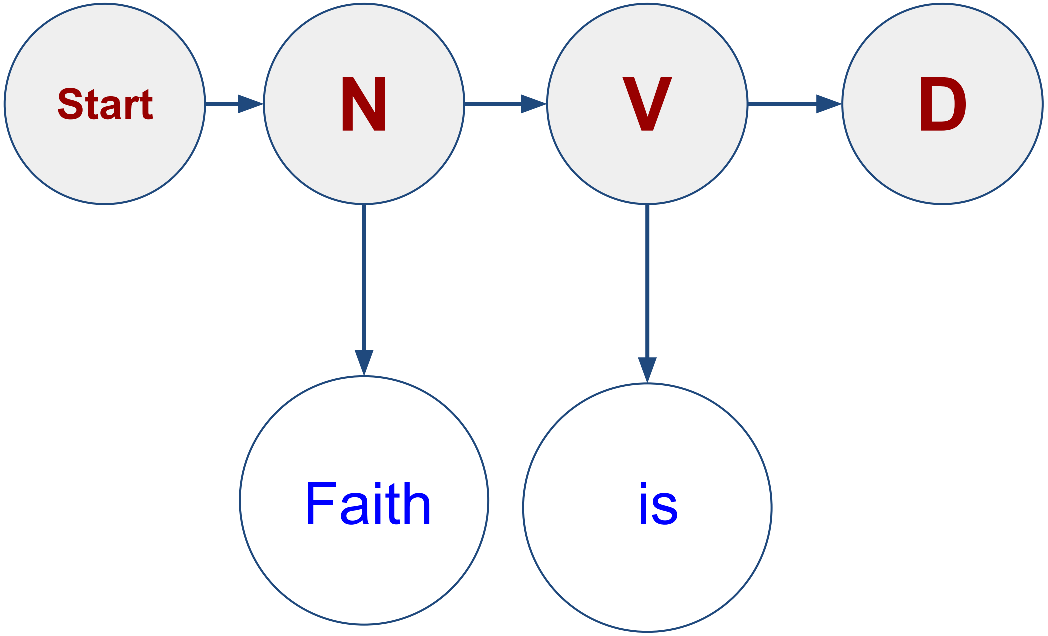
Sequence Labeling



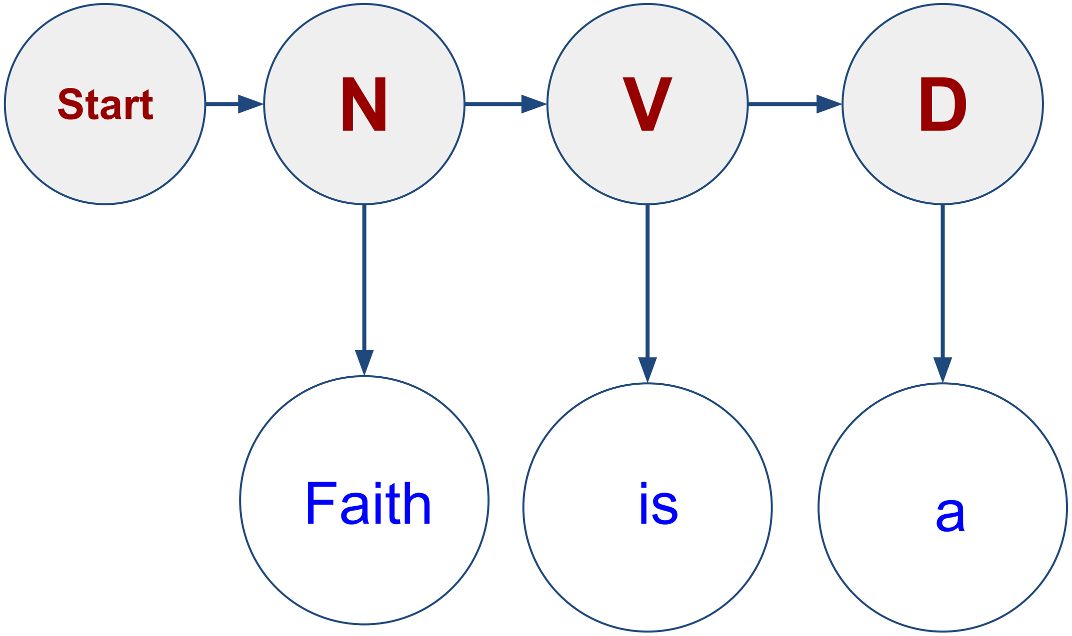
Sequence Labeling



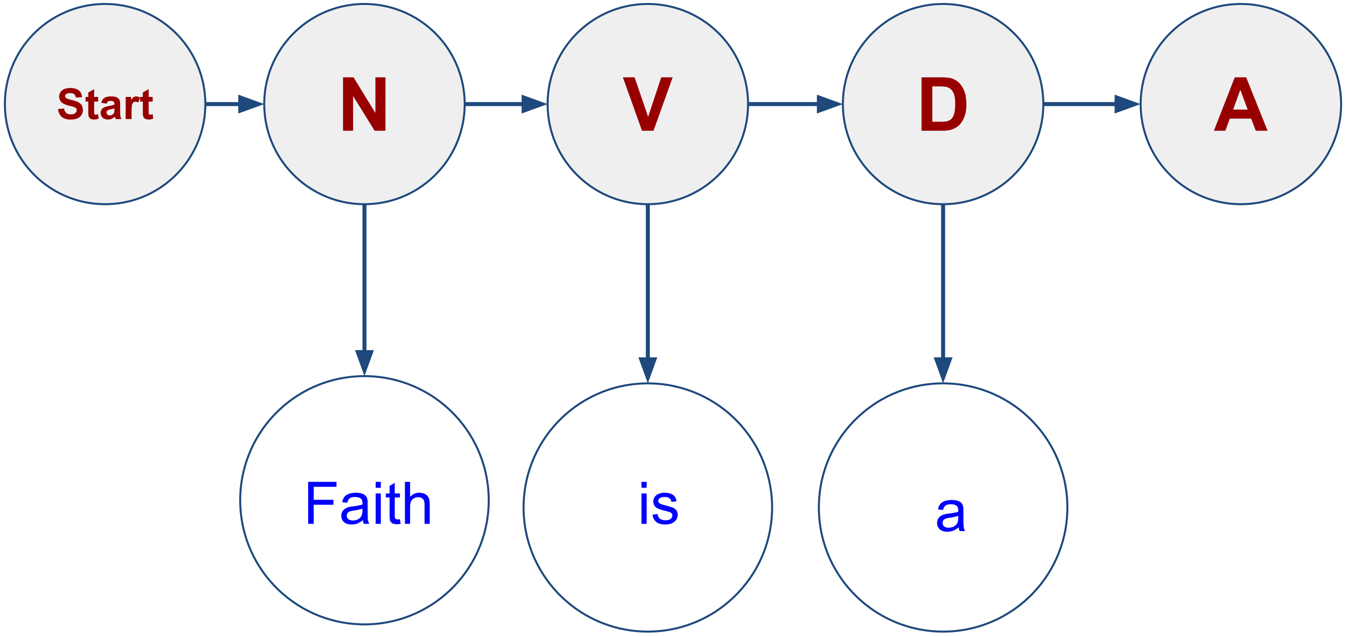
Sequence Labeling



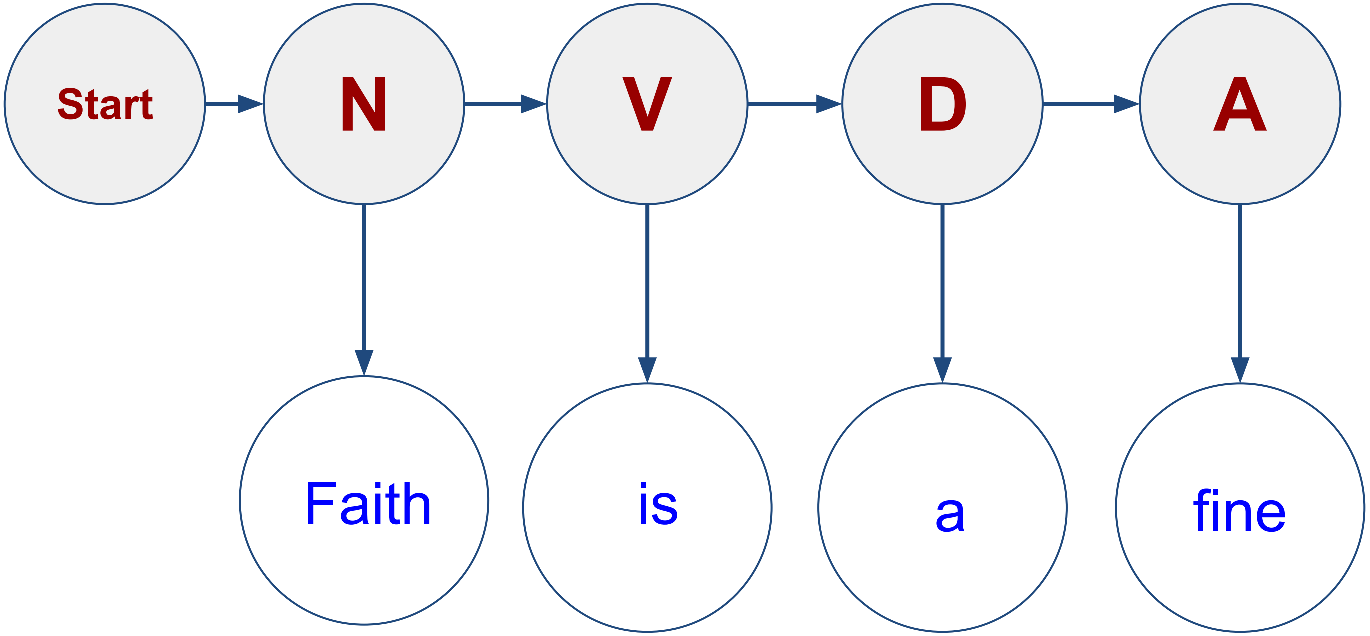
Sequence Labeling



Sequence Labeling

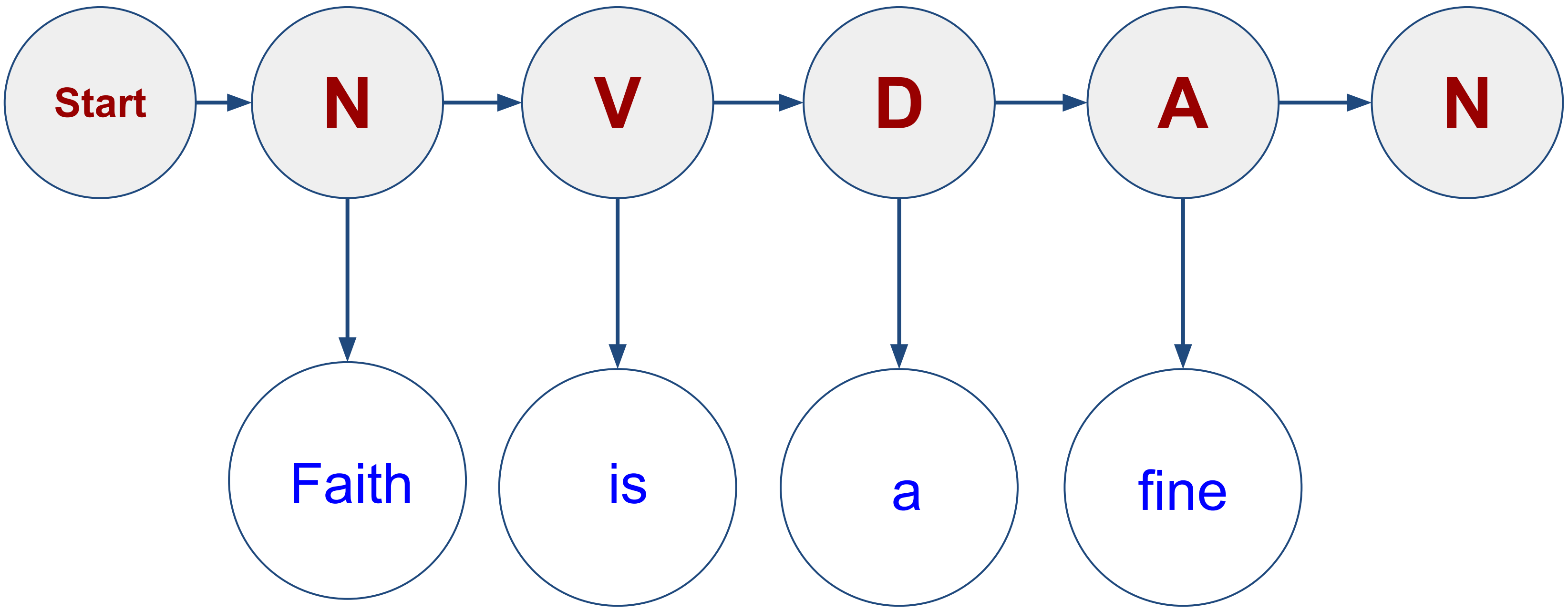


Sequence Labeling

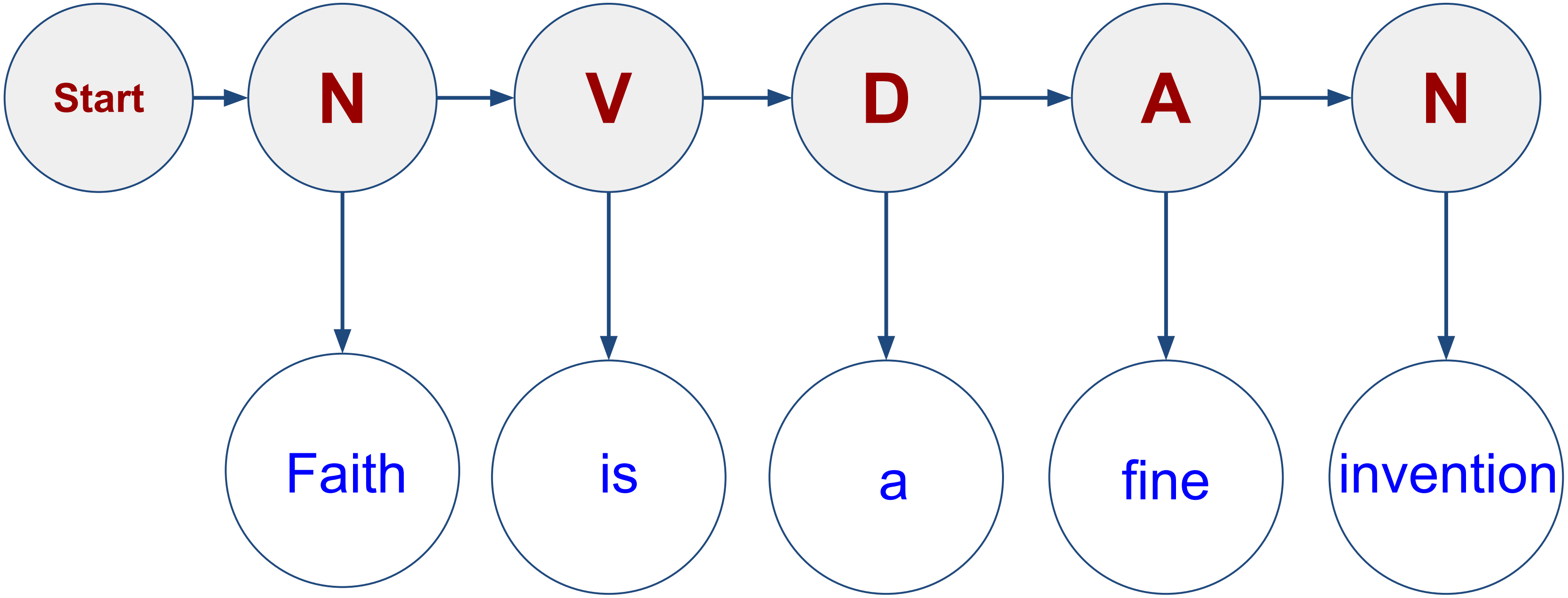


SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

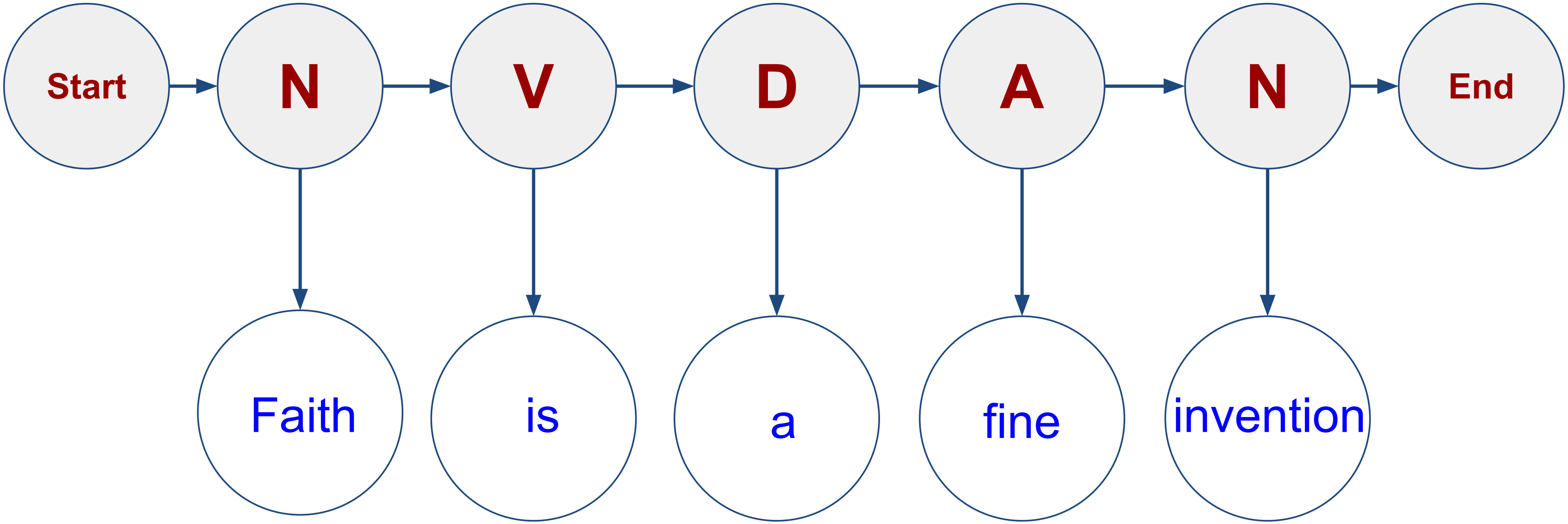
Sequence Labeling



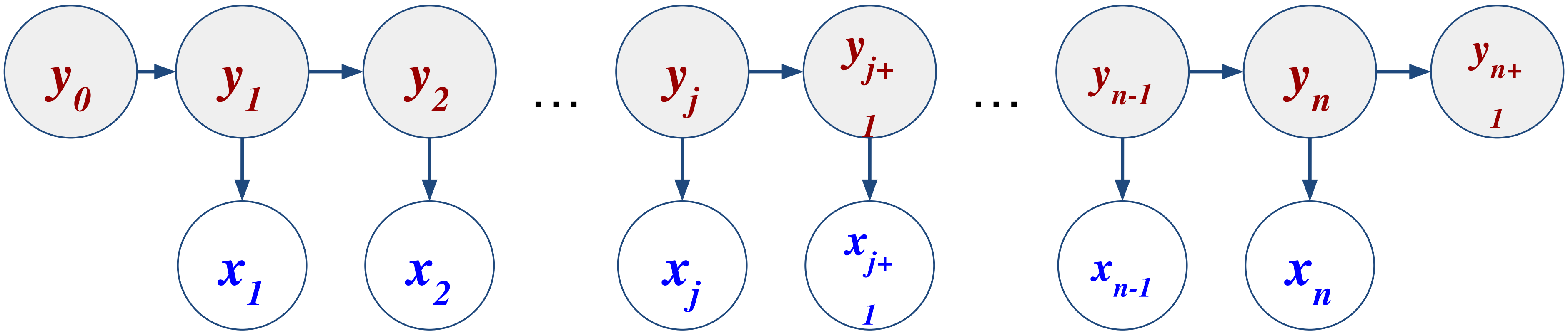
Sequence Labeling



Sequence Labeling



Sequence Labeling

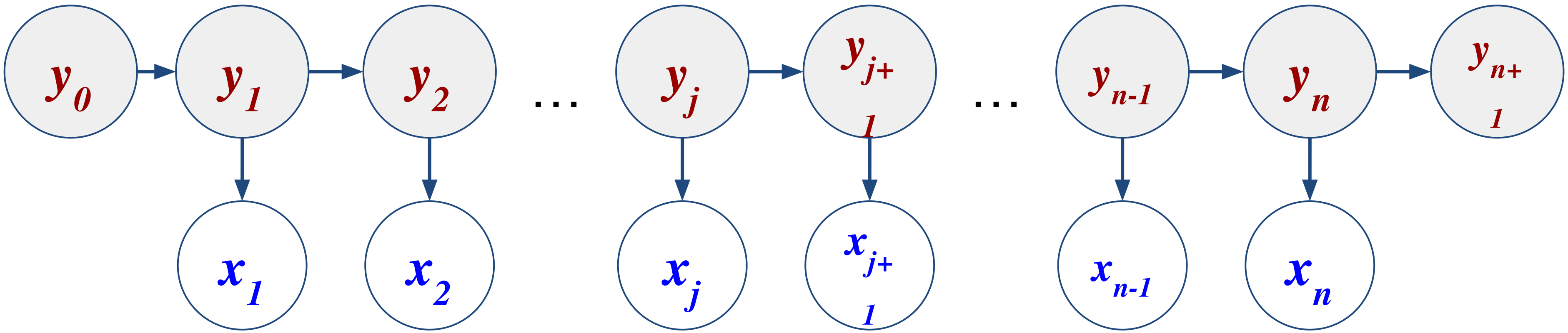


$$p(x_1, x_2, \ldots, x_{n-1}, x_n, y_0, y_1, y_2, \ldots, y_{n-1}, y_n, y_{n+1})$$

START

END

Sequence Labeling

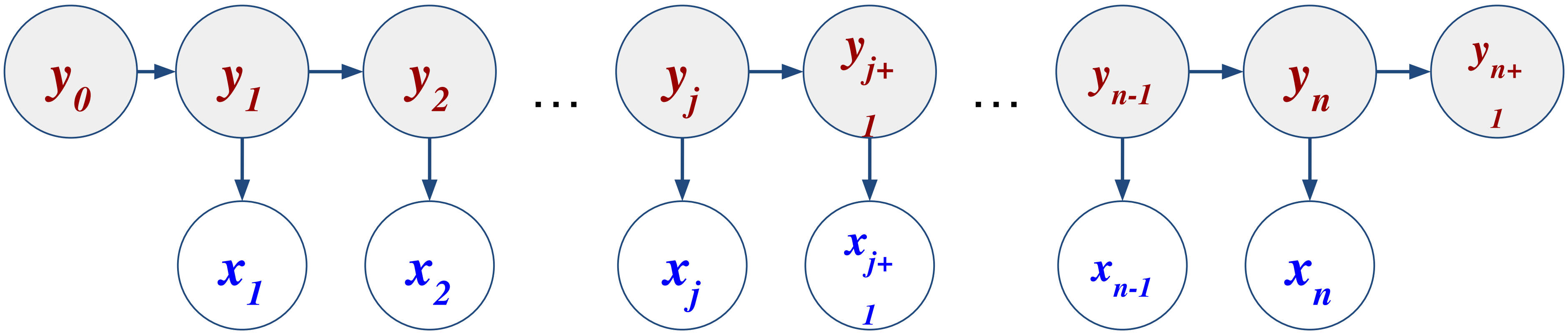


$$p(x_1, x_2, \ldots, x_{n-1}, x_n, y_0, y_1, y_2, \ldots, y_{n-1}, y_n, y_{n+1})$$

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

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Sequence Labeling

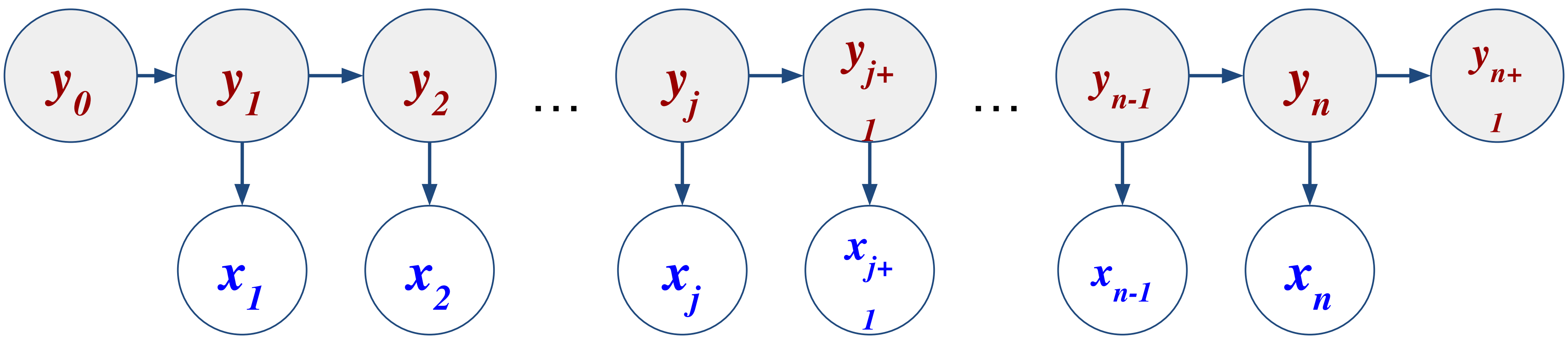


$$p(x_1, x_2, \ldots, x_{n-1}, x_n, y_0, y_1, y_2, \ldots, y_{n-1}, y_n, y_{n+1})$$

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

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Sequence Labeling



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

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

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 END states
  - $\mathcal{O}$: a set of observation symbols
  - θ : Transition and emission parameters $a_{u,v}$, and $b_u(o)$

Hidden Markov Model (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

$$a_{u,v}$$

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

$$b_u(o)$$

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Hidden Markov Model (Example)

$$a_{u,v}$$

$u \backslash v$	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

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

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

Hidden Markov Model (Example)

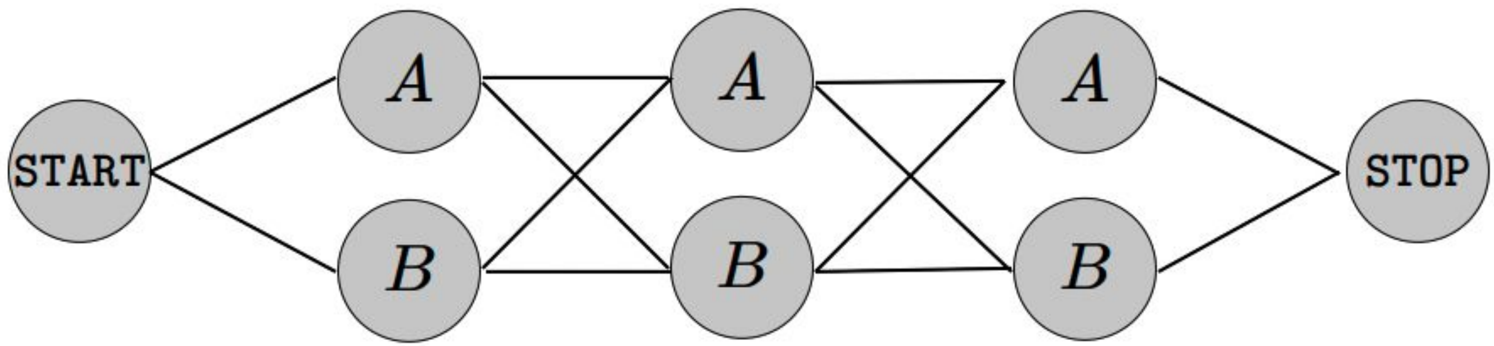
$a_{u,v}$

$u \backslash v$	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

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



SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Hidden Markov Model (Example)

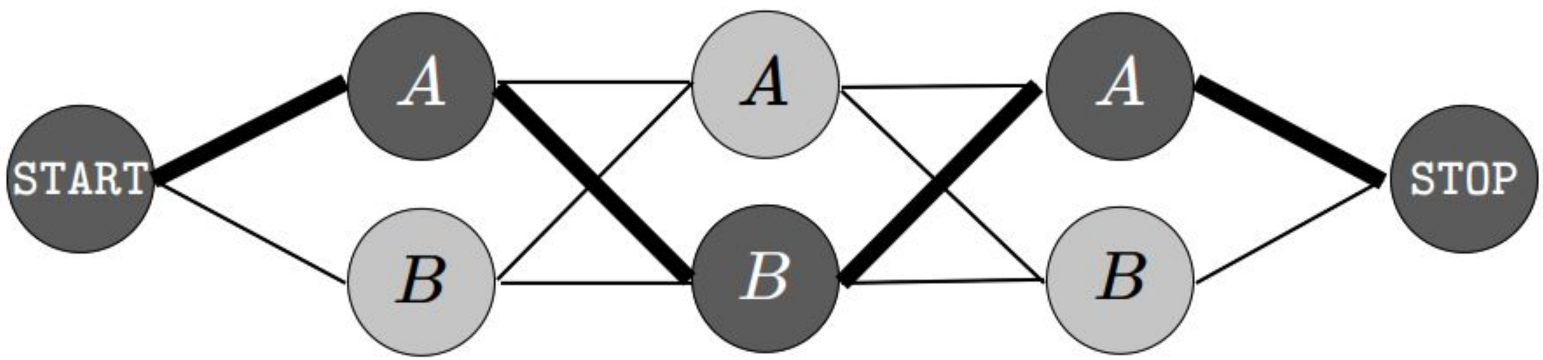
$a_{u,v}$

$u\backslash v$	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

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



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

Hidden Markov Model (Example)

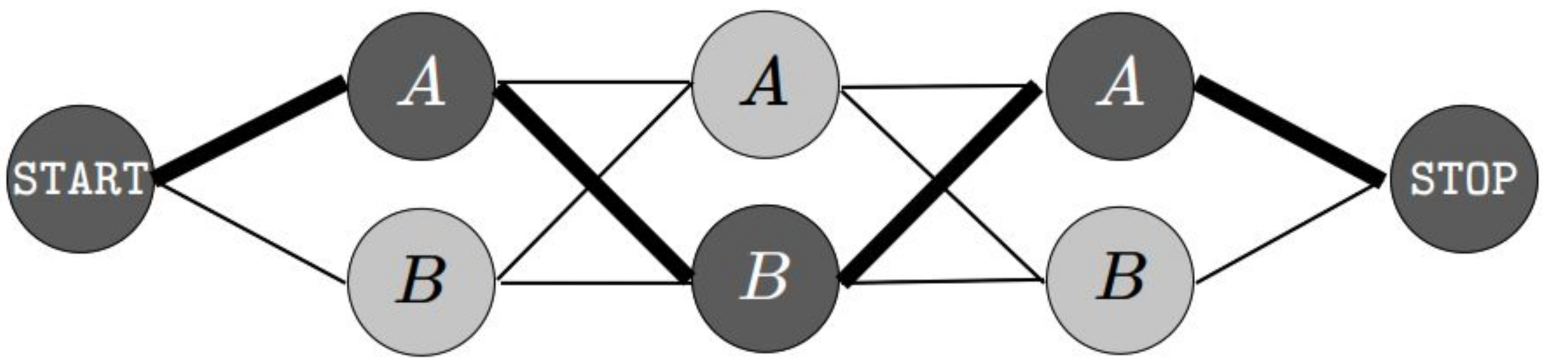
$a_{u,v}$

$u \backslash v$	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

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



$a_{\texttt{START},A} \times b_A(\text{"the"})$

Hidden Markov Model (Example)

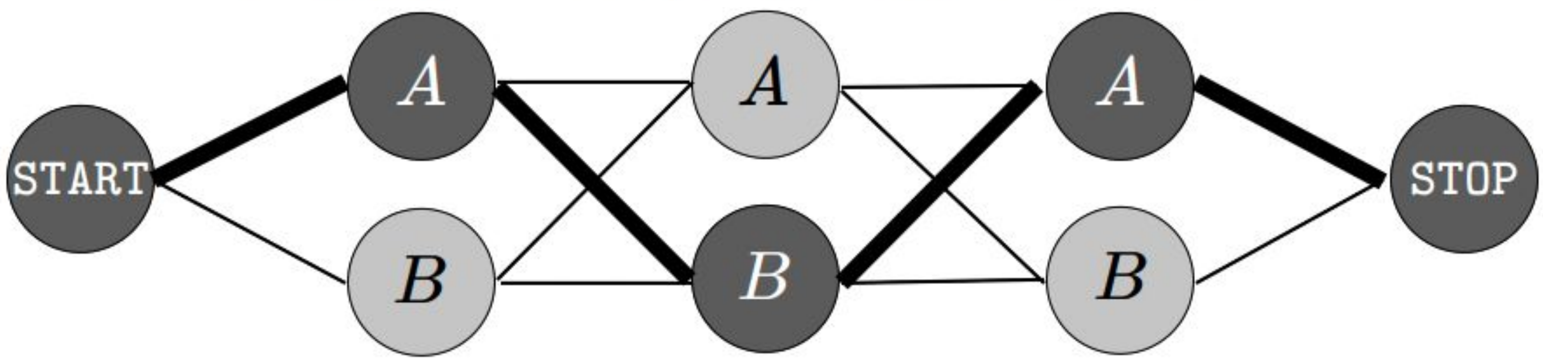
$a_{u,v}$

$u \backslash v$	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

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



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

Hidden Markov Model (Example)

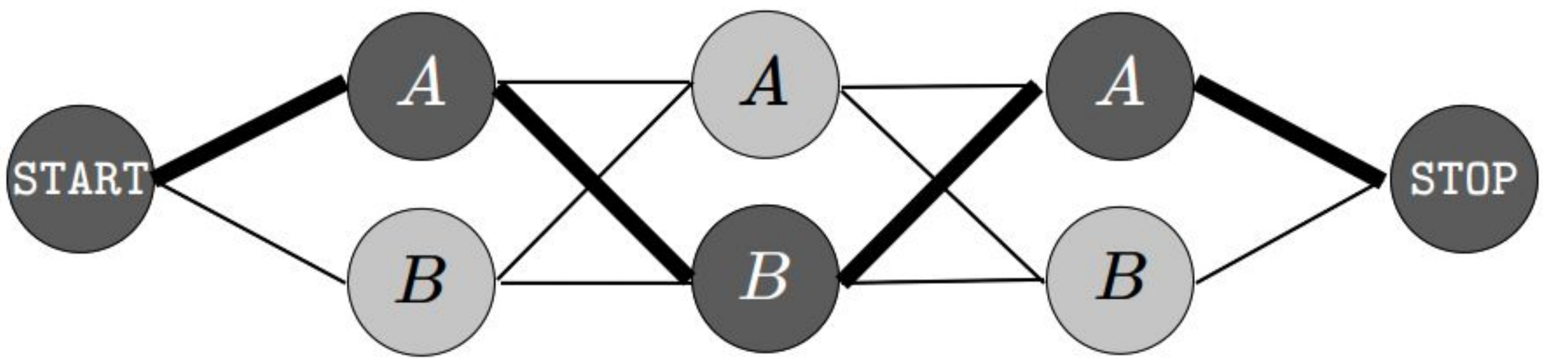
$$a_{u,v}$$

$u \backslash v$	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

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



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

Hidden Markov Model (Example)

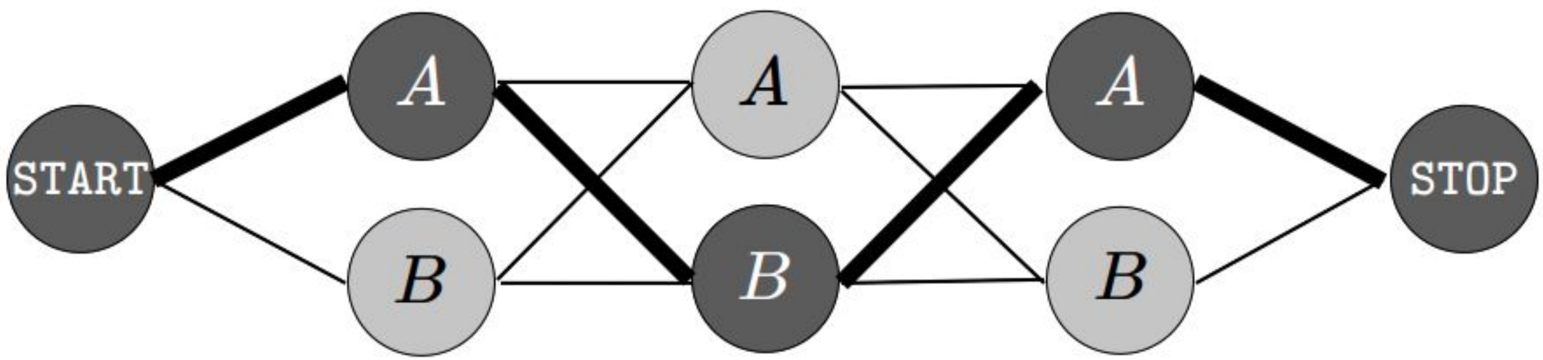
$a_{u,v}$

$u\backslash v$	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

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



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

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Hidden Markov Model (Example)

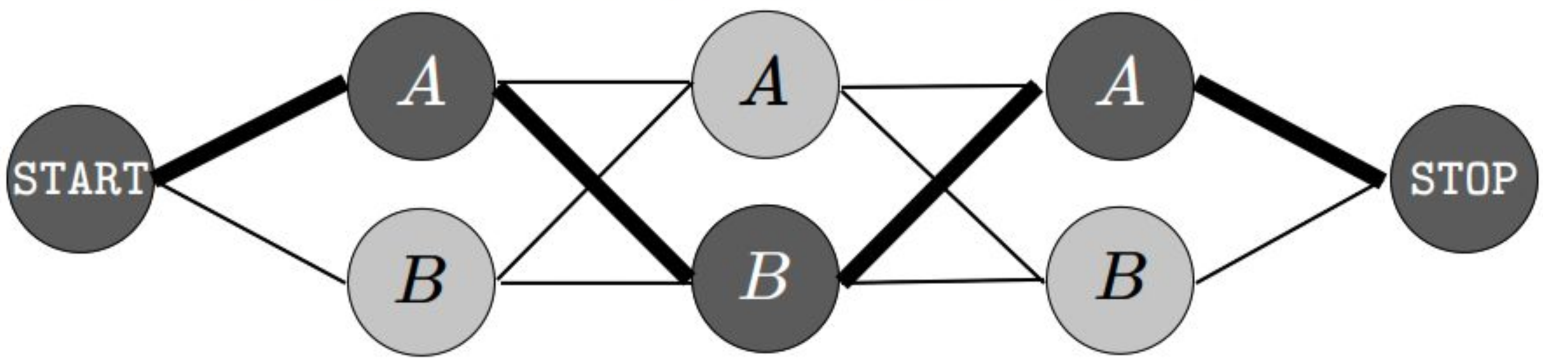
$a_{u,v}$

$u \backslash v$	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

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



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

Hidden Markov Model (Example)

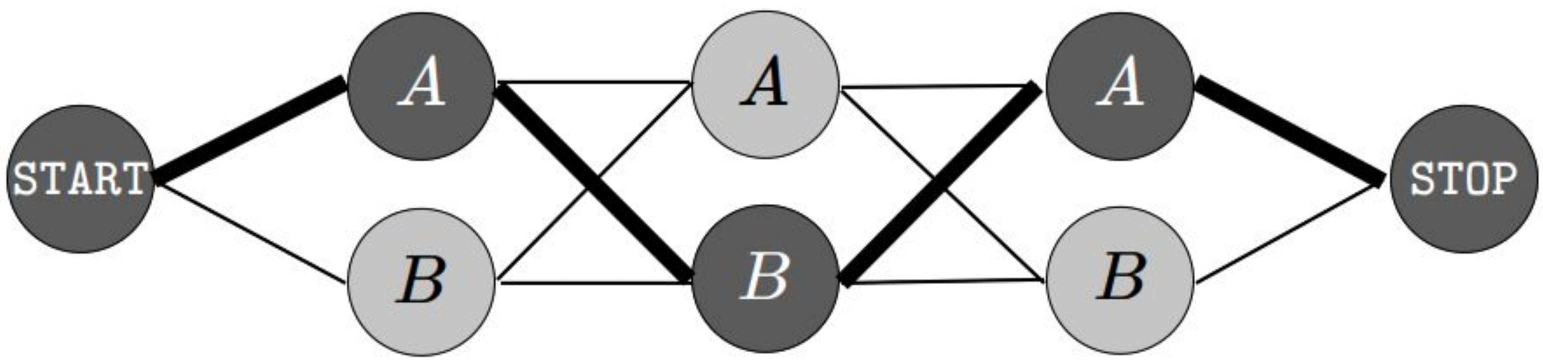
$$a_{u,v}$$

$u \backslash v$	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

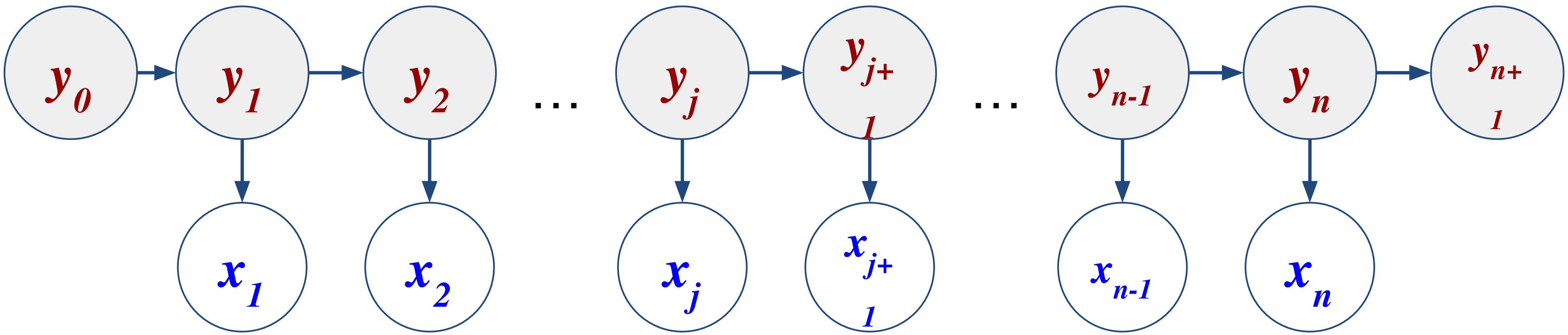
$$(\mathbf{x}, \mathbf{y}) = \text{the}/A, \text{dog}/B, \text{the}/A$$



$$a_{\texttt{START},A} \times b_A(\text{``the''}) \times a_{A,B} \times b_B(\text{``dog''}) \times a_{B,A} \times b_A(\text{``the''}) \times a_{A,\texttt{STOP}}$$

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

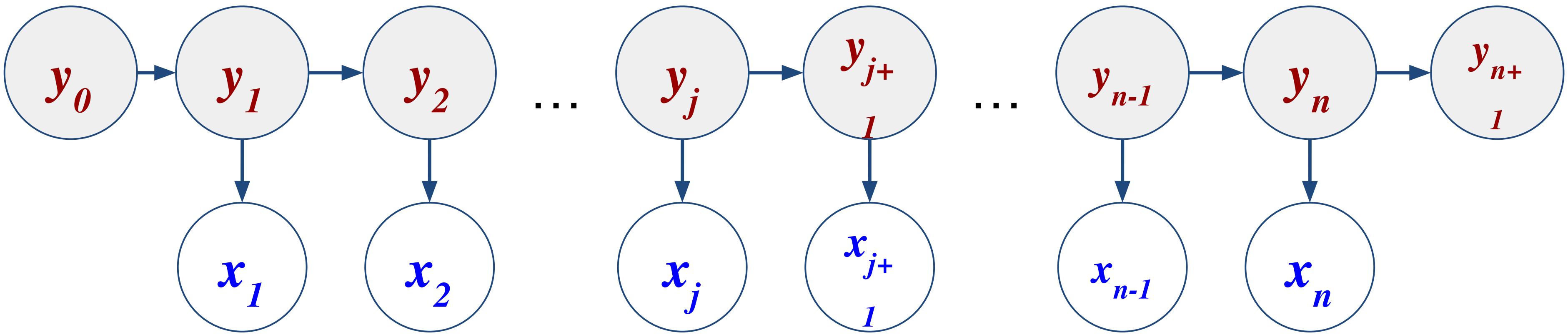
Hidden Markov Model



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

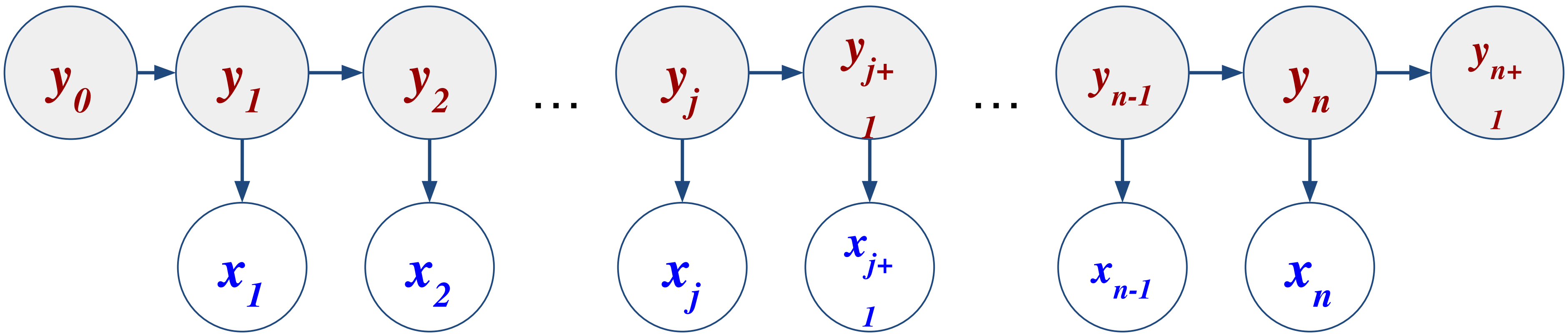
Now that we know what are the model parameters, how do we estimate them? In other words, how to do learning?

Hidden Markov Model



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

Hidden Markov Model



Number of times we see a transition from u to v

$$a_{u,v} = \frac{\boxed{\text{count}(u,v)}}{\boxed{\text{count}(u)}}$$

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

Number of times we see observation o generated from u

$$b_u(o) = \frac{\boxed{\text{count}(u \rightarrow o)}}{\boxed{\text{count}(u)}}$$

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

Hidden Markov Model

$$\begin{array}{ccc} y^{(1)} & = & (A, B) \\ x^{(1)} & & (e, g) \end{array} \qquad \begin{array}{ccc} y^{(2)} & = & (A, B) \\ x^{(2)} & & (e, h) \end{array} \qquad \begin{array}{ccc} y^{(3)} & = & (A, B) \\ x^{(3)} & & (f, h) \end{array}$$

$$a_{\mathtt{START},A} \times b_A(e) \times a_{A,B} \times b_B(g) \times a_{B,\mathtt{STOP}}$$

$$\times a_{\mathtt{START},A} \times b_A(e) \times a_{A,B} \times b_B(h) \times a_{B,\mathtt{STOP}}$$

$$\times a_{\mathtt{START},A} \times b_A(f) \times a_{A,B} \times b_B(h) \times a_{B,\mathtt{STOP}}$$

(1) Write down the likelihood

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Hidden Markov Model

$$\begin{array}{rcl} y^{(1)} & = & (A, B) \\ x^{(1)} & & (e, g) \end{array} \qquad \begin{array}{rcl} y^{(2)} & = & (A, B) \\ x^{(2)} & & (e, h) \end{array} \qquad \begin{array}{rcl} y^{(3)} & = & (A, B) \\ x^{(3)} & & (f, h) \end{array}$$

$$\boxed{a_{\mathtt{START},A}} \times \boxed{b_A(e)} \times \boxed{a_{A,B}} \times \boxed{b_B(g)} \times \boxed{a_{B,\mathtt{STOP}}}$$

$$\times \boxed{a_{\mathtt{START},A}} \times \boxed{b_A(e)} \times \boxed{a_{A,B}} \times \boxed{b_B(h)} \times \boxed{a_{B,\mathtt{STOP}}}$$

$$\times \boxed{a_{\mathtt{START},A}} \times \boxed{b_A(f)} \times \boxed{a_{A,B}} \times \boxed{b_B(h)} \times \boxed{a_{B,\mathtt{STOP}}}$$

Hidden Markov Model

$$\begin{array}{ccc} y^{(1)} & = & (A, B) \\ x^{(1)} & & (e, g) \end{array} \qquad \begin{array}{ccc} y^{(2)} & = & (A, B) \\ x^{(2)} & & (e, h) \end{array} \qquad \begin{array}{ccc} y^{(3)} & = & (A, B) \\ x^{(3)} & & (f, h) \end{array}$$

$$\left(\boxed{a_{\texttt{START},A}}\right)^3 \times \left(\boxed{a_{A,B}}\right)^3 \times \left(\boxed{a_{B,\texttt{STOP}}}\right)^3$$
$$\times \left(\boxed{b_A(f)}\right)^1 \times \left(\boxed{b_A(e)}\right)^2$$
$$\times \left(\boxed{b_B(h)}\right)^2 \times \left(\boxed{b_B(g)}\right)^1$$

Hidden Markov Model

$$\begin{array}{ccc} y^{(1)} & & (A, B) \\ x^{(1)} & = & (e, g) \end{array} \qquad \begin{array}{ccc} y^{(2)} & & (A, B) \\ x^{(2)} & = & (e, h) \end{array} \qquad \begin{array}{ccc} y^{(3)} & & (A, B) \\ x^{(3)} & = & (f, h) \end{array}$$

$$(a_{\texttt{START},A})^3 \times (a_{A,B})^3 \times (a_{B,\texttt{STOP}})^3$$

$$\times (b_A(f))^1 \times (b_A(e))^2$$

$$\times (b_B(h))^2 \times (b_B(g))^1$$

Hidden Markov Model

$$\begin{array}{rcl} y^{(1)} & \multirow{2}{*}{=} & (A, B) \\ x^{(1)} & & (e, g) \end{array} \qquad \begin{array}{rcl} y^{(2)} & \multirow{2}{*}{=} & (A, B) \\ x^{(2)} & & (e, h) \end{array} \qquad \begin{array}{rcl} y^{(3)} & \multirow{2}{*}{=} & (A, B) \\ x^{(3)} & & (f, h) \end{array}$$

$$\begin{aligned} & \left(a_{\mathtt{START},A}\right)^{\mathrm{Count}(\mathtt{START},A)} \\ \times & \left(a_{A,B}\right)^{\mathrm{Count}(A,B)} \\ \times & \left(a_{B,\mathtt{STOP}}\right)^{\mathrm{Count}(B,\mathtt{STOP})} \end{aligned}$$

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Hidden Markov Model

$$\begin{array}{ccc} y^{(1)} & & (A, B) \\ x^{(1)} & = & (e, g) \end{array} \qquad \begin{array}{ccc} y^{(2)} & & (A, B) \\ x^{(2)} & = & (e, h) \end{array} \qquad \begin{array}{ccc} y^{(3)} & & (A, B) \\ x^{(3)} & = & (f, h) \end{array}$$

$$\prod_{u,v} (a_{u,v})^{\mathrm{Count}(u,v)}$$

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Hidden Markov Model

$$\begin{matrix} y^{(1)} \\ x^{(1)} \end{matrix} = \begin{matrix} (A, B) \\ (e, g) \end{matrix} \qquad \begin{matrix} y^{(2)} \\ x^{(2)} \end{matrix} = \begin{matrix} (A, B) \\ (e, h) \end{matrix} \qquad \begin{matrix} y^{(3)} \\ x^{(3)} \end{matrix} = \begin{matrix} (A, B) \\ (f, h) \end{matrix}$$

$$\prod_{u,v} (a_{u,v})^{\mathrm{Count}(u,v)}$$

$$\times \prod_{u,o} (b_u(o))^{\mathrm{Count}(u \to o)}$$

(2) This is the likelihood. We can consider the log-likelihood.

Hidden Markov Model

$$\begin{array}{ccc} y^{(1)} \\ x^{(1)} \end{array} = \begin{array}{c} (A, B) \\ (e, g) \end{array} \qquad \begin{array}{c} y^{(2)} \\ x^{(2)} \end{array} = \begin{array}{c} (A, B) \\ (e, h) \end{array} \qquad \begin{array}{c} y^{(3)} \\ x^{(3)} \end{array} = \begin{array}{c} (A, B) \\ (f, h) \end{array}$$

$$\sum_{u,v} \mathrm{Count}(u, v) \log(a_{u,v})$$

$$+ \sum_{u,o} \mathrm{Count}(u \rightarrow o) \log(b_u(o))$$

(3) Next, take the partial derivative with respect to each individual parameter, and set it to zero. Solve the equation.

Hidden Markov Model

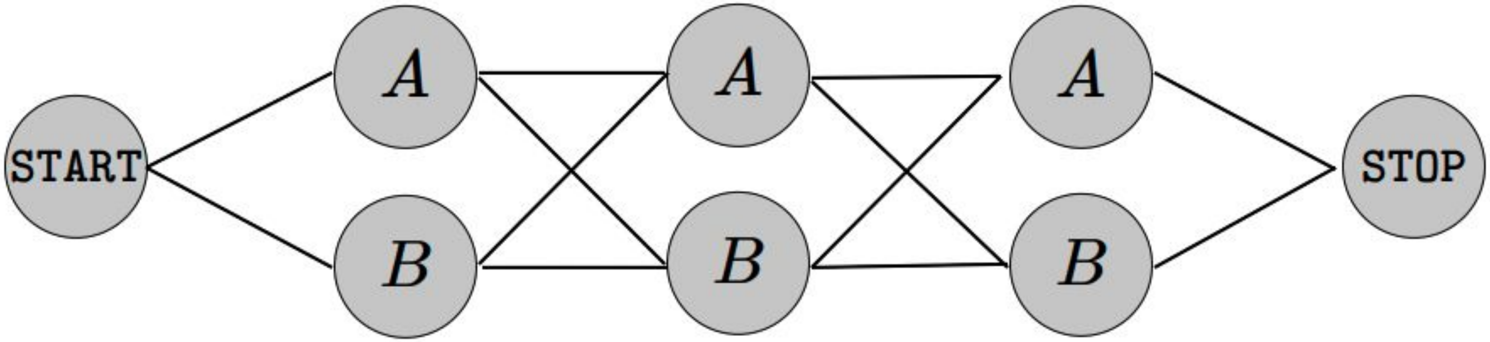
$a_{u,v}$

$u\backslash v$	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

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



Which label sequence y is the most probable given the word sequence x?