Solutions for Assignment 3

Exercise 1 - Bigram Inference

You are given the following training corpus:

1. `<s> I am Sam </s>`
2. `<s> Sam I am </s>`
3. `<s> Sam I like </s>`
4. `<s> Sam I do like </s>`
5. `<s> do I like Sam </s>`

Assume now that you have trained a bigram language model on this corpus.

Exercise 1 – Building the model

- Probability of a sequence of words is given by:

$$P(W = \vec{w}) = P(w_1) * P(w_2|w_1) * P(w_3|w_1, w_2) * ... * P(w_n|w_1, ..., w_{n-1})$$

=> Next word in seq. depends on all previous words

- Modeling quickly becomes intractable!
- Solution - Markov assumption (bigram):

$$P(w_i|w_1, ..., w_{i-1}) = P(w_i|w_{i-1})$$

=> Next word in seq. only depends on previous word

$$P(W = \vec{w}) = P(w_1) * P(w_2|w_1) * P(w_3|w_2) * ... * P(w_n|w_{n-1})$$

Exercise 1 – Building the model

- Bigram-model approximates probabilities using MLE of a corpus:

$$P(w_i \mid w_{i-1}) = \frac{count(w_{i-1}, w_i)}{count(w_{i-1})}$$

- Probability of next word in the sequence given current word is count of the pair normalized by count of the current word
- Counts are computed using a corpus of sequences
- Convention: Sequences always start with „<s>“: $P(w_1 =$„<s>“) = 1

Exercise 1 – Building the model

- Example:

Current word w_{i-1} = „am“

1. <s> I am Sam </s>
2. <s> Sam I am </s>
3. <s> Sam I like </s>
4. <s> Sam I do like </s>
5. <s> do I like Sam </s>

- count(w_{i-1} =„am“) = 2
- count((w_{i-1}, w_i) =(„am“, „Sam“)) = 1

=> $P(w_i$ =„Sam“| w_{i-1} =„am“) = $\frac{1}{2}$

- count((w_{i-1}, w_i) =(„am“, „</s>“)) = 1

=> $P(w_i$ =„</s>“| w_{i-1} =„am“) = $\frac{1}{2}$

- $P(w_i$ =„sth. else“| w_{i-1} =„am“) = 0

Exercise 1 – Building the model

- For convenience: Build a lookup table:

W_{i-1}/W_i	<s>	am	do	I	like	Sam	</s>
<s>	0	0	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{3}{5}$	0
am	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
do	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
I	0	$\frac{2}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}$	0	0
like	0	0	0	0	0	$\frac{1}{3}$	$\frac{2}{3}$
Sam	0	0	0	$\frac{3}{5}$	0	0	$\frac{2}{5}$

Our example

1. What is the most probable next word predicted by the model for the following word sequences?

   (a) `<s> Sam ...`

   (b) `<s> Sam I do ...`

   (c) `<s> Sam I am Sam ...`

   (d) `<s> do I like ...`

Exercise 1 – 1.a)

(a) <s> Sam . . . → w_{i-1} = „Sam“

W_{i-1}/W_i	<s>	am	do	I	like	Sam	</s>
<s>	0	0	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{3}{5}$	0
am	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
do	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
I	0	$\frac{2}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}$	0	0
like	0	0	0	0	0	$\frac{1}{3}$	$\frac{2}{3}$
Sam	0	0	0	$\frac{3}{5}$	0	0	$\frac{2}{5}$

$P(w_i = „\text{I}“ | w_{i-1} = „\text{Sam}“) = \frac{3}{5}$

$P(w_i = „\text{</s>}“ | w_{i-1} = „\text{Sam}“) = \frac{2}{5}$

$P(w_i = \text{other} | w_{i-1} = „\text{Sam}“) = 0$

Most probable next word:
„I“

Exercise 1 – 2.

1. What is the most probable next word predicted by the model for the following word sequences?

   (a) `<s> Sam ...` → „I“

   (b) `<s> Sam I do ...` → „like“ or „I“

   (c) `<s> Sam I am Sam ...` → „I“

   (d) `<s> do I like ...` → „</s>“

2. Which of the following sentences is better, i.e., gets a higher probability with this model?

   (a) `<s> Sam I am </s>`
   (b) `<s> Sam I do I like </s>`
   (c) `<s>I do like Sam I am</s>`

Exercise 1 – 2.a)

(a) <s> Sam I am </s>

$$P(W) = P(\underset{w_1}{\text{"<s>"}}, \text{"Sam"}, \text{"I"}, \text{"am"}, \text{"Sam"}, \text{"</s>"})$$

$$= P(\text{„<s>“})$$

Exercise 1 – 2.a)

(a) <s> Sam I am </s>

$$P(W) = P(\text{"}\underset{\substack{\uparrow\\ w_{i-1}}}{\text{<s>}}\text{", "}\underset{\substack{\uparrow\\ w_i}}{\text{Sam}}\text{", "I", "am", "Sam", "</s>"})$$

$$= P(\text{„<s>“})P(\text{„Sam“}|\text{„<s>“})$$

Exercise 1 – 2.a)

(a) <s> Sam I am </s>

$$P(W) = P(\text{"<s>"}, \underset{\underset{w_{i-1}}{\uparrow}}{\text{"Sam"}}, \underset{\underset{w_i}{\uparrow}}{\text{"I"}}, \text{"am"}, \text{"Sam"}, \text{"</s>"})$$

$$= P(\text{„<s>“})P(\text{„Sam“}|\text{„<s>“})\, P(\text{„I“}|\text{„Sam“})$$

Exercise 1 – 2.a)

(a) <s> Sam I am </s>

$P(W)$ =P("<s>", "Sam", "I", "am", "Sam", "</s>")

w_{i-1} w_i

= P(„<s>“)P(„Sam“|„<s>“) P(„I“|„Sam“)P(„am“|„I“)

Exercise 1 – 2.a)

(a) <s> Sam I am </s>

$$P(W) = P(\text{"<s>"}, \text{"Sam"}, \text{"I"}, \text{"am"}, \text{"Sam"}, \text{"</s>"})$$

w_{i-1} w_i

$$= P(\text{„<s>“})P(\text{„Sam“}|\text{„<s>“})\,P(\text{„I“}|\text{„Sam“})P(\text{„am“}|\text{„I“})P(\text{„Sam“}|\text{„am“})$$

Exercise 1 – 2.a)

(a) <s> Sam I am </s>

$$P(W) = P(\text{"<s>"}, \text{"Sam"}, \text{"I"}, \text{"am"}, \underset{\underset{w_{i-1}}{\uparrow}}{\text{"Sam"}}, \underset{\underset{w_i}{\uparrow}}{\text{"</s>"}})$$

$$= P(\text{„<s>“})P(\text{„Sam“}|\text{„<s>“})\,P(\text{„I“}|\text{„Sam“})P(\text{„am“}|\text{„I“})P(\text{„Sam“}|\text{„am“})$$
$$P(\text{„</s>“}|\text{„Sam“})$$

Exercise 1 – 2.a)

$P(W) = P(„\text{<s>}“)P(„\text{Sam}“|„\text{<s>}“)P(„\text{I}“|„\text{Sam}“)P(„\text{am}“|„\text{I}“)P(„\text{Sam}“|„\text{am}“)P(„\text{</s>}“|„\text{Sam}“)$

$= \quad 1$

W_{i-1}/W_i	<s>	am	do	I	like	Sam	</s>
<s>	0	0	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{3}{5}$	0
am	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
do	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
I	0	$\frac{2}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}$	0	0
like	0	0	0	0	0	$\frac{1}{3}$	$\frac{2}{3}$
Sam	0	0	0	$\frac{3}{5}$	0	0	$\frac{2}{5}$

Exercise 1 – 2.a)

$$P(W) = P(\text{„<s>“})P(\text{„Sam“}|\text{„<s>“})P(\text{„I“}|\text{„Sam“})P(\text{„am“}|\text{„I“})P(\text{„Sam“}|\text{„am“})P(\text{„</s>“}|\text{„Sam“})$$

$$= \quad 1 \quad * \quad \frac{3}{5}$$

W_{i-1}/W_i	<s>	am	do	I	like	Sam	</s>
<s>	0	0	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{3}{5}$	0
am	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
do	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
I	0	$\frac{2}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}$	0	0
like	0	0	0	0	0	$\frac{1}{3}$	$\frac{2}{3}$
Sam	0	0	0	$\frac{3}{5}$	0	0	$\frac{2}{5}$

Exercise 1 – 2.a)

$P(W) = P(„\text{<s>}“)P(„\text{Sam}“|„\text{<s>}“)P(„\text{I}“|„\text{Sam}“)P(„\text{am}“|„\text{I}“)P(„\text{Sam}“|„\text{am}“)P(„\text{</s>}“|„\text{Sam}“)$

$= 1 * \frac{3}{5} * \frac{3}{5}$

W_{i-1}/W_i	<s>	am	do	I	like	Sam	</s>
<s>	0	0	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{3}{5}$	0
am	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
do	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
I	0	$\frac{2}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}$	0	0
like	0	0	0	0	0	$\frac{1}{3}$	$\frac{2}{3}$
Sam	0	0	0	$\frac{3}{5}$	0	0	$\frac{2}{5}$

Exercise 1 – 2.a)

$$P(W) = P(\text{„<s>“})P(\text{„Sam“}|\text{„<s>“})P(\text{„I“}|\text{„Sam“})P(\text{„am“}|\text{„I“})P(\text{„Sam“}|\text{„am“})P(\text{„</s>“}|\text{„Sam“})$$

$$= 1 \quad * \quad \frac{3}{5} \quad * \quad \frac{3}{5} \quad * \quad \frac{2}{5}$$

W_{i-1}/W_i	<s>	am	do	I	like	Sam	</s>
<s>	0	0	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{3}{5}$	0
am	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
do	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
I	0	$\frac{2}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}$	0	0
like	0	0	0	0	0	$\frac{1}{3}$	$\frac{2}{3}$
Sam	0	0	0	$\frac{3}{5}$	0	0	$\frac{2}{5}$

Exercise 1 – 2.a)

$P(W) = P(„\text{<s>}“)P(„\text{Sam}“|„\text{<s>}“)P(„\text{I}“|„\text{Sam}“)P(„\text{am}“|„\text{I}“)P(„\text{Sam}“|„\text{am}“)P(„\text{</s>}“|„\text{Sam}“)$

$= 1 * \frac{3}{5} * \frac{3}{5} * \frac{2}{5} * \frac{1}{2}$

W_{i-1}/W_i	<s>	am	do	I	like	Sam	</s>
<s>	0	0	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{3}{5}$	0
am	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
do	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
I	0	$\frac{2}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}$	0	0
like	0	0	0	0	0	$\frac{1}{3}$	$\frac{2}{3}$
Sam	0	0	0	$\frac{3}{5}$	0	0	$\frac{2}{5}$

Exercise 1 – 2.a)

$$P(W) = P(\text{„<s>“})P(\text{„Sam“}|\text{„<s>“})P(\text{„I“}|\text{„Sam“})P(\text{„am“}|\text{„I“})P(\text{„Sam“}|\text{„am“})P(\text{„</s>“}|\text{„Sam“})$$

$$= 1 * \frac{3}{5} * \frac{3}{5} * \frac{2}{5} * \frac{1}{2} * \frac{2}{5}$$

W_{i-1}/W_i	<s>	am	do	I	like	Sam	</s>
<s>	0	0	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{3}{5}$	0
am	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
do	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
I	0	$\frac{2}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}$	0	0
like	0	0	0	0	0	$\frac{1}{3}$	$\frac{2}{3}$
Sam	0	0	0	$\frac{3}{5}$	0	0	$\frac{2}{5}$

Exercise 1 – 2.a)

$$P(W) = P(\text{„<s>“})P(\text{„Sam“}|\text{„<s>“})P(\text{„I“}|\text{„Sam“})P(\text{„am“}|\text{„I“})P(\text{„Sam“}|\text{„am“})P(\text{„</s>“}|\text{„Sam“})$$

$$= 1 * \frac{3}{5} * \frac{3}{5} * \frac{2}{5} * \frac{1}{2} * \frac{2}{5}$$

$$= \frac{36}{1250}$$

$$= 2.88\%$$

Exercise 1 – 2.

2. Which of the following sentences is better, i.e., gets a higher probability with this model?

   (a) `<s> Sam I am </s>` → 2.88% best sequence

   (b) `<s> Sam I do I like </s>` → 0.96%

   (c) `<s>I do like Sam I am</s>` → 0.08%

3. Compute the perplexity of the model for the following sequence (note that, in general, start-of-sentence tokens are excluded when calculating perplexity):
`<s> I do like Sam`

Exercise 1 – 3.

- Perplexity in general:

$$PP(W) = \sqrt[N]{\frac{1}{P(w_1 w_2 \ldots w_N)}}$$

- Inverse likelihood of sequence under the model normalized by sequence length

Exercise 1 – 3.

- Perplexity in a bigram model:

$$PP(W) = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(w_i|w_{i-1})}}$$

Exercise 1 – 3.

- Likelihood of given sequence:

$$P(\text{W="<s>", "I", "do", "like", "Sam"}) = 1 * \frac{1}{5} * \frac{1}{5} * \frac{1}{2} * \frac{1}{3}$$

$$= \frac{1}{150}$$

- Perplexity:

$$PP(W) = \sqrt[4]{\frac{1}{\frac{1}{150}}}$$

$$= \sqrt[4]{150}$$

Exercise 2 - Character recognition using HMM

Given the structure of hidden states (see figure 1) and the learned HMM for character 'A' and the learned HMM for character 'B' as follows:

$$A^{(\text{letter A})} = \begin{bmatrix} .8 & .2 & 0 \\ 0 & .8 & .2 \\ 0 & 0 & 1 \end{bmatrix} \; B^{(\text{letter A})} = \begin{bmatrix} .9 & .1 & 0 \\ .1 & .8 & .1 \\ .9 & .1 & 0 \end{bmatrix}$$

and similarly for letter "B":

$$A^{(\text{letter B})} = \begin{bmatrix} .8 & .2 & 0 \\ 0 & .8 & .2 \\ 0 & 0 & 1 \end{bmatrix} \; B^{(\text{letter B})} = \begin{bmatrix} .9 & .1 & 0 \\ 0 & .2 & .8 \\ .6 & .4 & 0 \end{bmatrix}$$

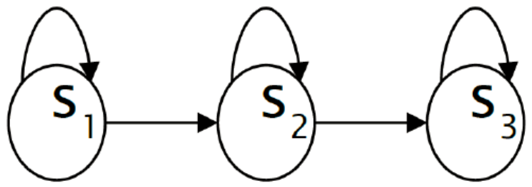


Figure 1: Structure of the hidden states

Suppose that after character image segmentation the following sequence of island numbers in 4 slices was observed (see figure 2):

$$\vec{o} = (1, 3, 2, 1)$$

What HMM is more likely to generate this observation sequence, HMM for 'A' or HMM for 'B'?

Figure 2: An example of a vertical slice for both characters.

Exercise 2 – possible state transitions

For our example there are a total of 27 possible state transition sequences:

1. s1 -> s1 -> s1 -> s1
2. s1 -> s1 -> s1 -> s2

...

27. s1 -> s3 -> s3 -> s3

Exercise 2 -Filtering

We can filter:

- „Backtracking“ sequences are impossible (i.e. s2 -> s1)
- „Skips“ are impossible (i.e. s1 -> s3)
- Observing 3 islands is impossible in s1 and s3
- Assertion: No state may repeat more than once

=> Only 2 sequences are possible under both models:

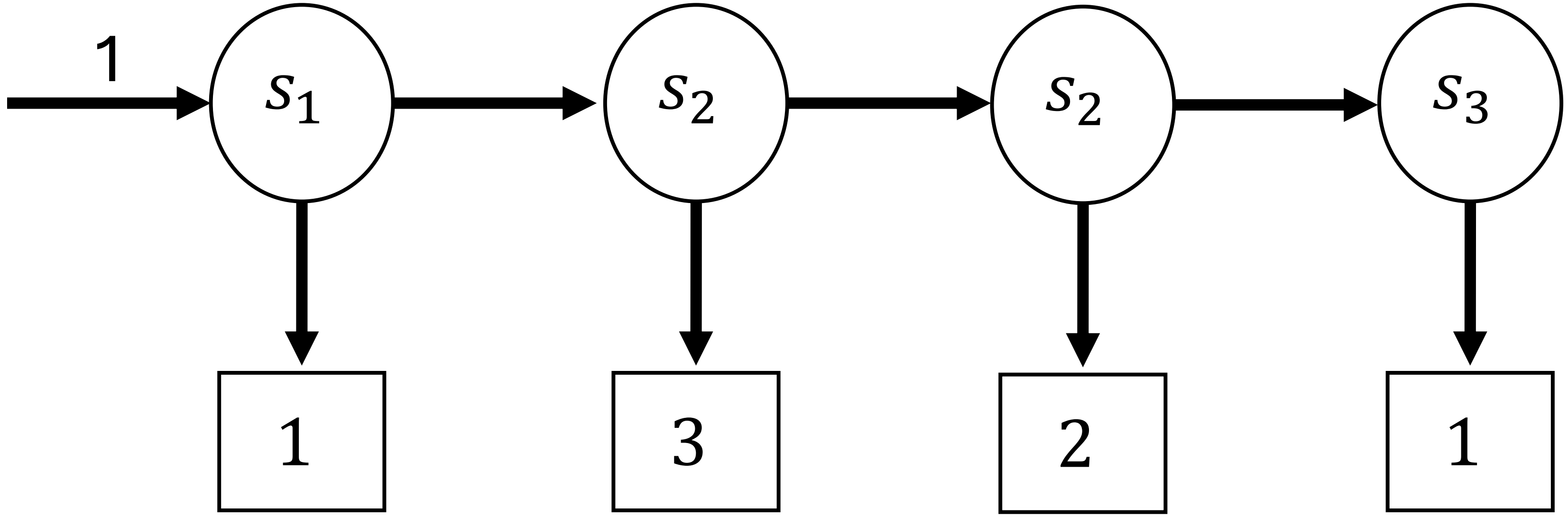
1. s1 -> s2 -> s2 -> s3
2. s1 -> s2 -> s3 -> s3

=> Probability of $\vec{o}$: Sum of the likelihoods of these 2 sequences

Exercise 2 – Model A

Given:

- $\vec{s}_1 = (s_1, s_2, s_2, s_3)$
- $\vec{o} = (1,3,2,1)$

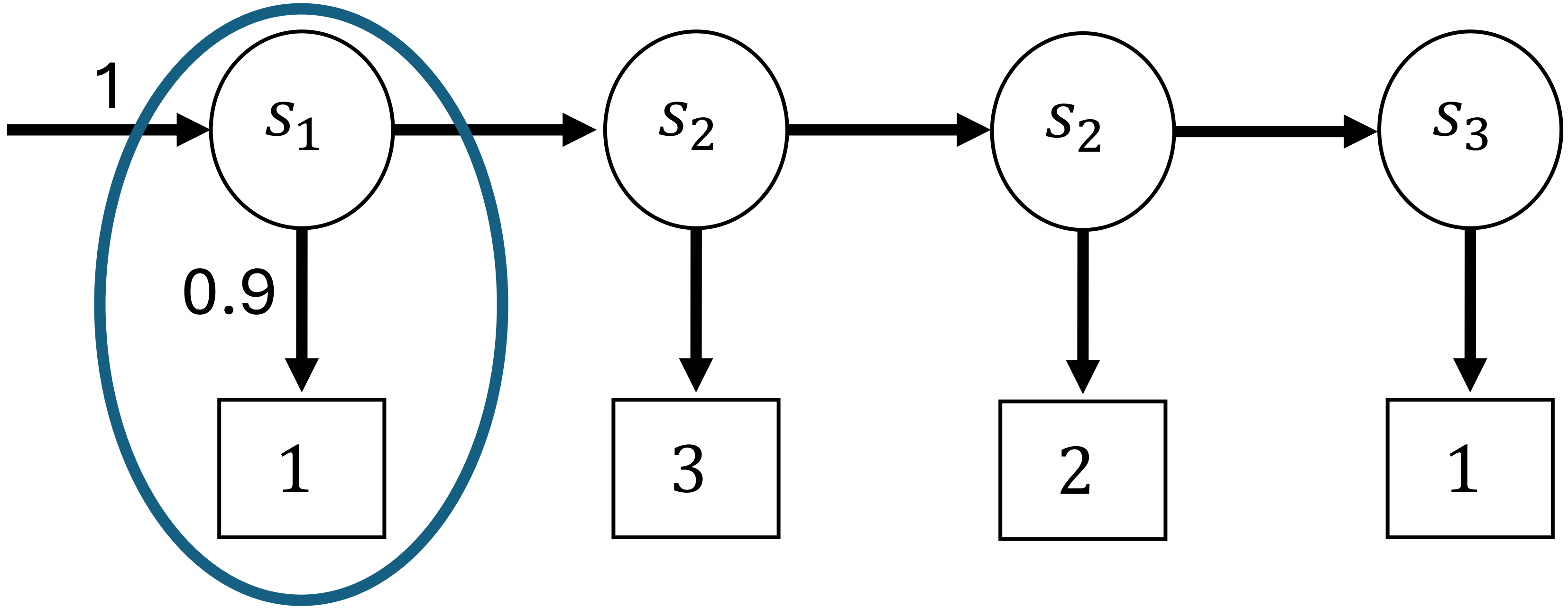


Exercise 2 – Model A

Given:

- $\vec{s}_1 = (s_1, s_2, s_2, s_3)$
- $\vec{o} = (1,3,2,1)$

$$A^{(\text{letter A})} = \begin{bmatrix} .8 & .2 & 0 \\ 0 & .8 & .2 \\ 0 & 0 & 1 \end{bmatrix} \quad B^{(\text{letter A})} = \begin{bmatrix} .9 & .1 & 0 \\ .1 & .8 & .1 \\ .9 & .1 & 0 \end{bmatrix}$$

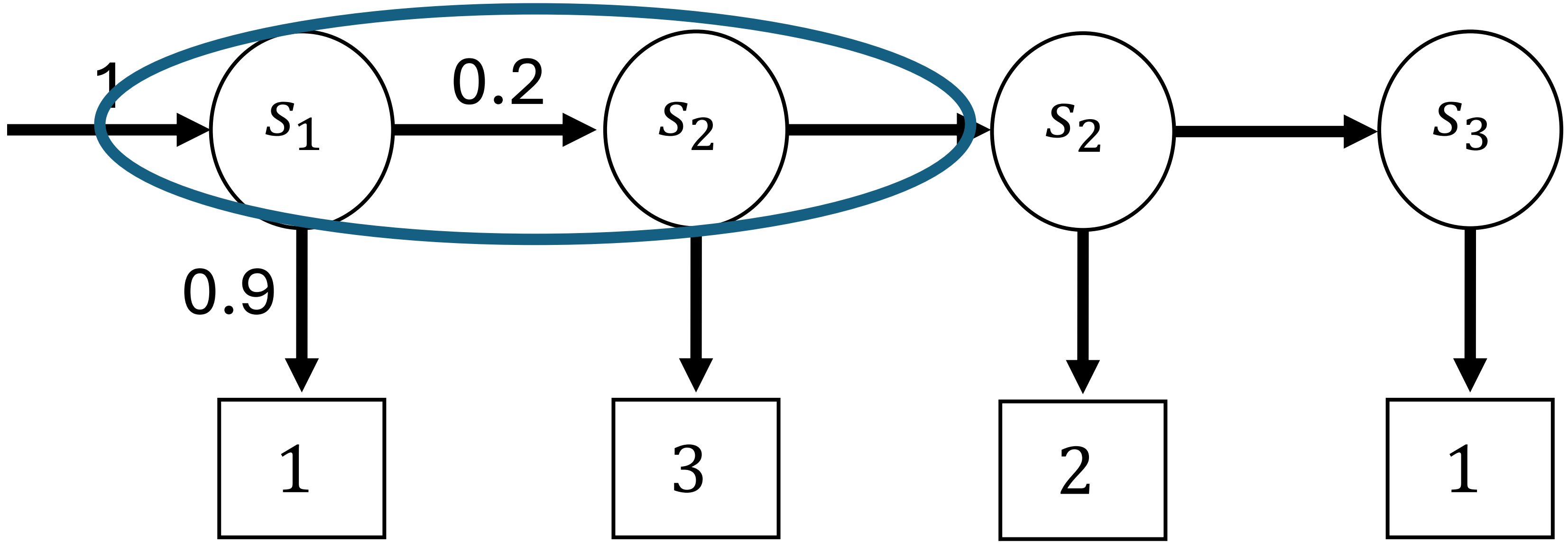


Exercise 2 – Model A

Given:

- $\vec{s}_1 = (s_1, s_2, s_2, s_3)$
- $\vec{o} = (1,3,2,1)$

$$A^{(\text{letter A})} = \begin{bmatrix} .8 & .2 & 0 \\ 0 & .8 & .2 \\ 0 & 0 & 1 \end{bmatrix} \quad B^{(\text{letter A})} = \begin{bmatrix} .9 & .1 & 0 \\ .1 & .8 & .1 \\ .9 & .1 & 0 \end{bmatrix}$$

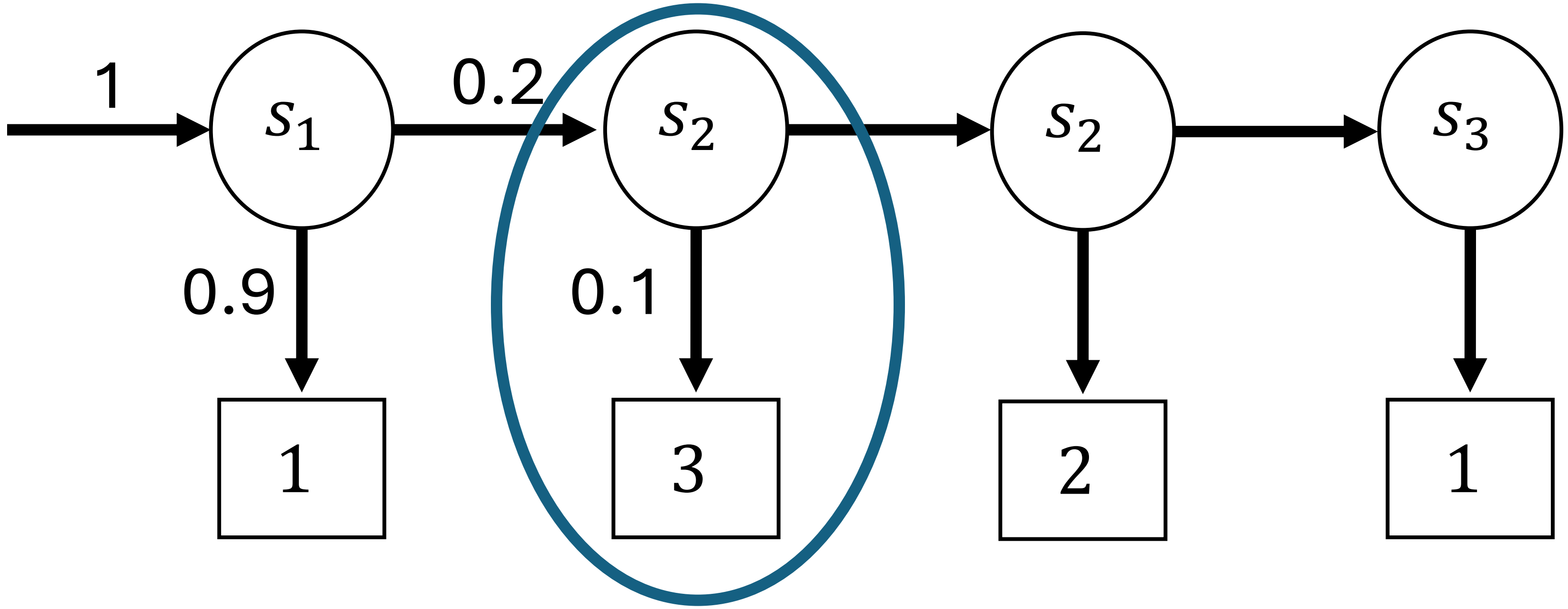


Exercise 2 – Model A

Given:

- $\vec{s}_1 = (s_1, s_2, s_2, s_3)$
- $\vec{o} = (1,3,2,1)$

$$A^{(\text{letter A})} = \begin{bmatrix} .8 & .2 & 0 \\ 0 & .8 & .2 \\ 0 & 0 & 1 \end{bmatrix} \quad B^{(\text{letter A})} = \begin{bmatrix} .9 & .1 & 0 \\ .1 & .8 & .1 \\ .9 & .1 & 0 \end{bmatrix}$$

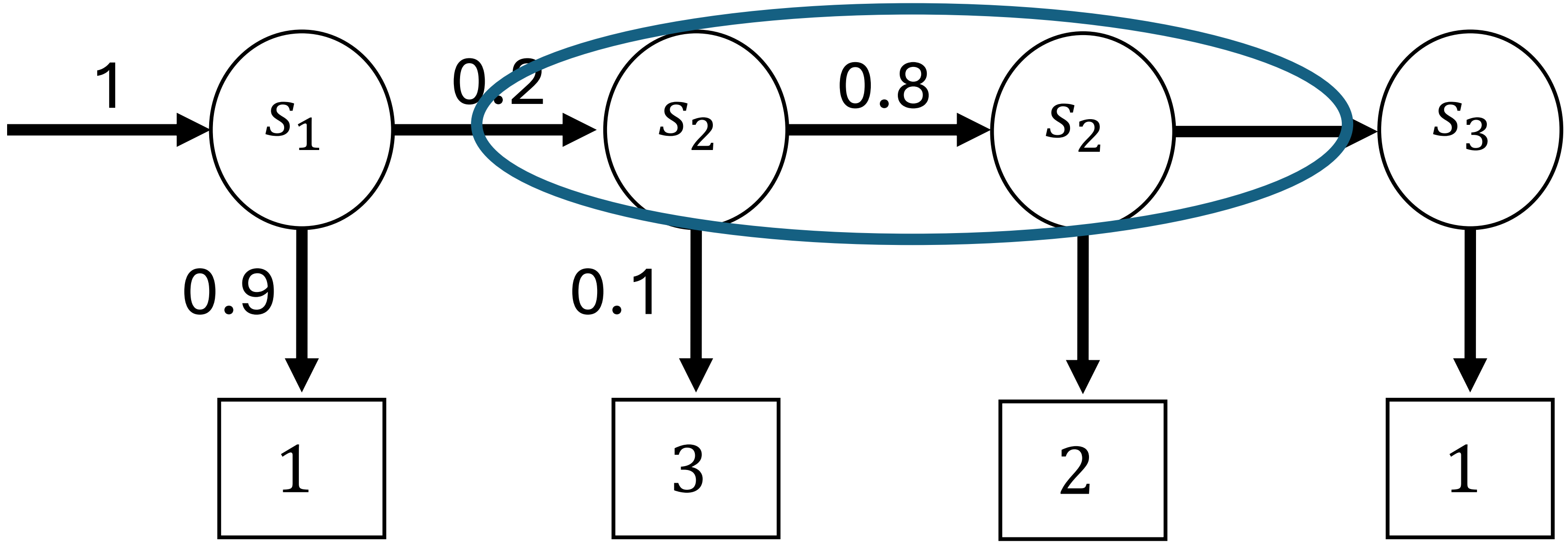


Exercise 2 – Model A

Given:

- $\vec{s}_1 = (s_1, s_2, s_2, s_3)$
- $\vec{o} = (1,3,2,1)$

$$A^{(\text{letter A})} = \begin{bmatrix} .8 & .2 & 0 \\ 0 & .8 & .2 \\ 0 & 0 & 1 \end{bmatrix} \quad B^{(\text{letter A})} = \begin{bmatrix} .9 & .1 & 0 \\ .1 & .8 & .1 \\ .9 & .1 & 0 \end{bmatrix}$$

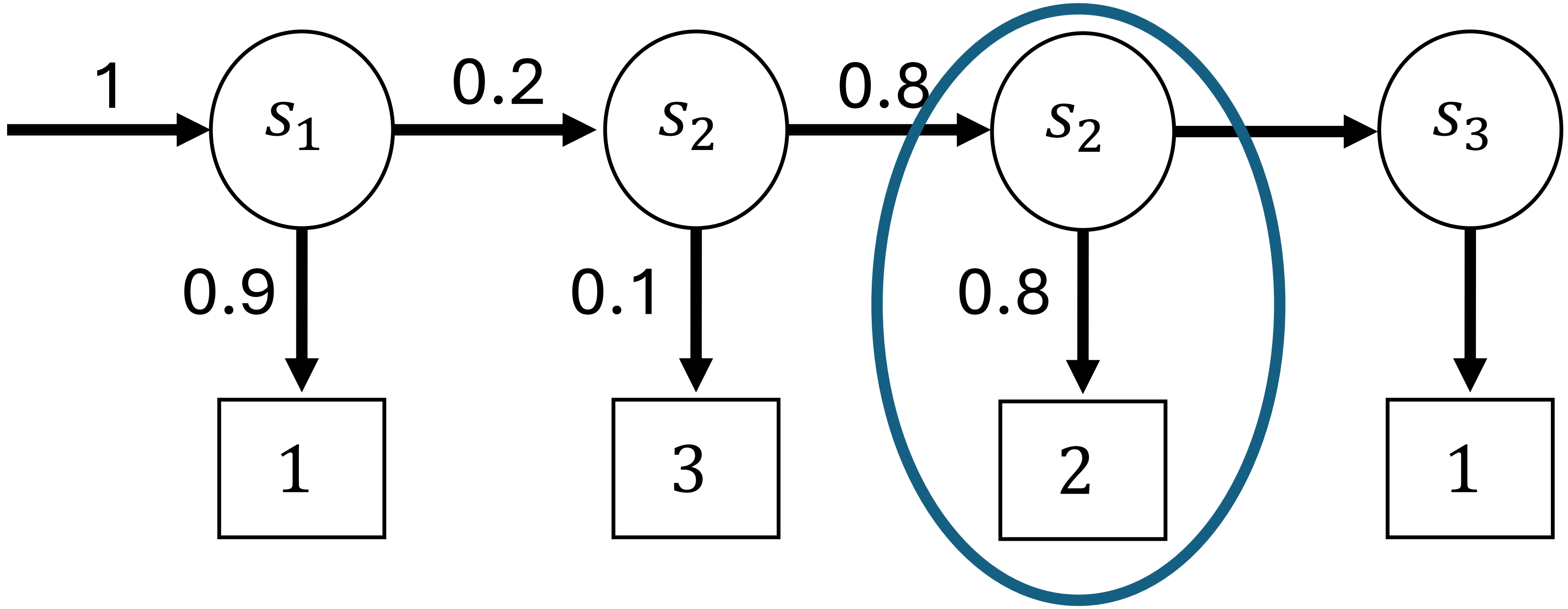


Exercise 2 – Model A

Given:

- $\vec{s}_1 = (s_1, s_2, s_2, s_3)$
- $\vec{o} = (1,3,2,1)$

$$A^{(\text{letter A})} = \begin{bmatrix} .8 & .2 & 0 \\ 0 & .8 & .2 \\ 0 & 0 & 1 \end{bmatrix} \quad B^{(\text{letter A})} = \begin{bmatrix} .9 & .1 & 0 \\ .1 & .8 & .1 \\ .9 & .1 & 0 \end{bmatrix}$$

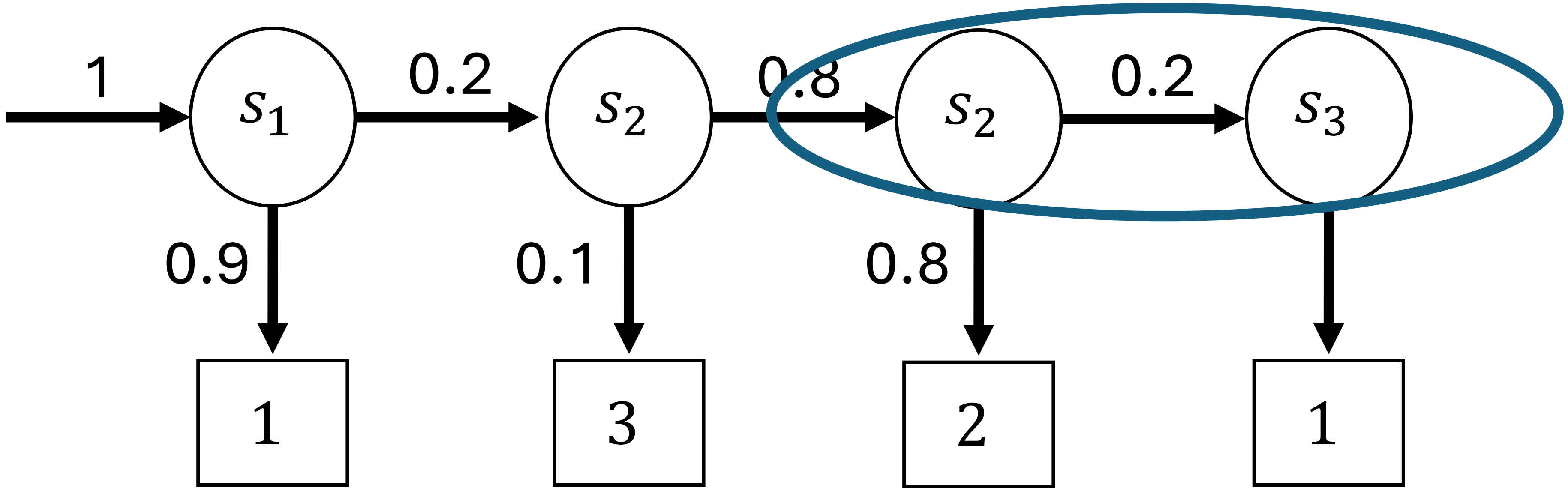


Exercise 2 – Model A

Given:

- $\vec{s}_1 = (s_1, s_2, s_2, s_3)$
- $\vec{o} = (1,3,2,1)$

$$A^{(\text{letter A})} = \begin{bmatrix} .8 & .2 & 0 \\ 0 & .8 & .2 \\ 0 & 0 & 1 \end{bmatrix} \quad B^{(\text{letter A})} = \begin{bmatrix} .9 & .1 & 0 \\ .1 & .8 & .1 \\ .9 & .1 & 0 \end{bmatrix}$$

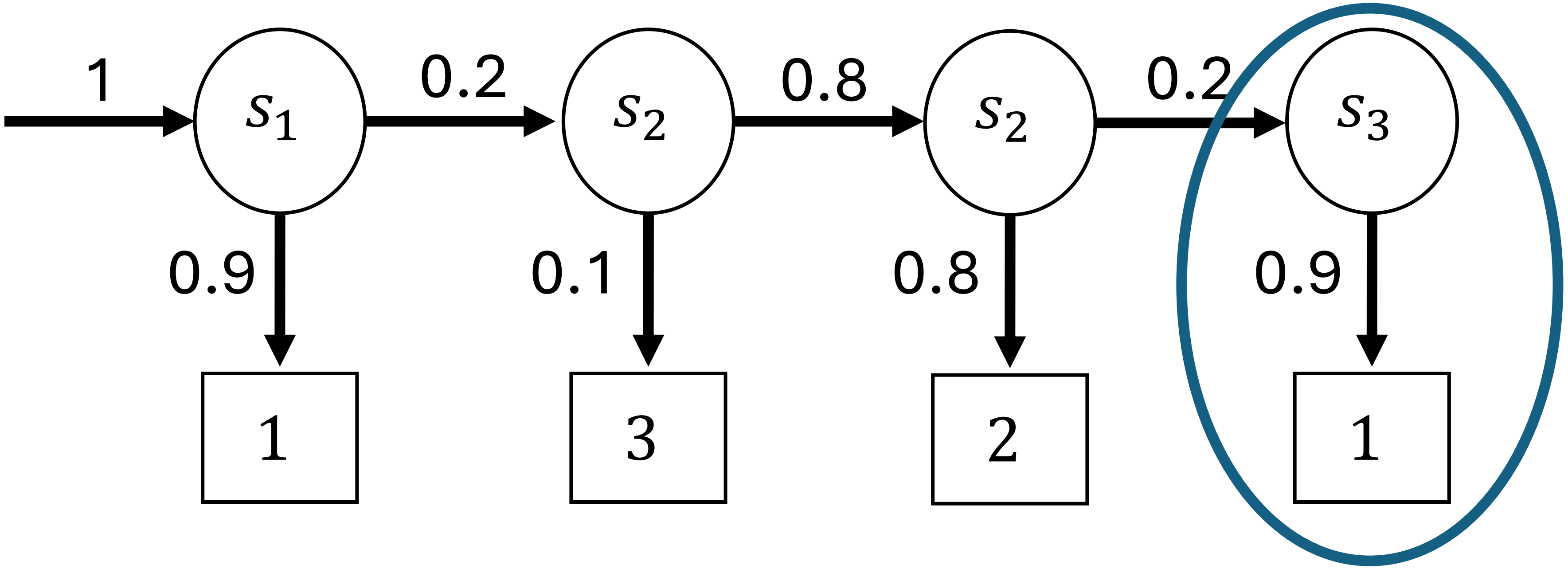


Exercise 2 – Model A

Given:

- $\vec{s}_1 = (s_1, s_2, s_2, s_3)$
- $\vec{o} = (1,3,2,1)$

$$A^{(\text{letter A})} = \begin{bmatrix} .8 & .2 & 0 \\ 0 & .8 & .2 \\ 0 & 0 & 1 \end{bmatrix} \quad B^{(\text{letter A})} = \begin{bmatrix} .9 & .1 & 0 \\ .1 & .8 & .1 \\ .9 & .1 & 0 \end{bmatrix}$$

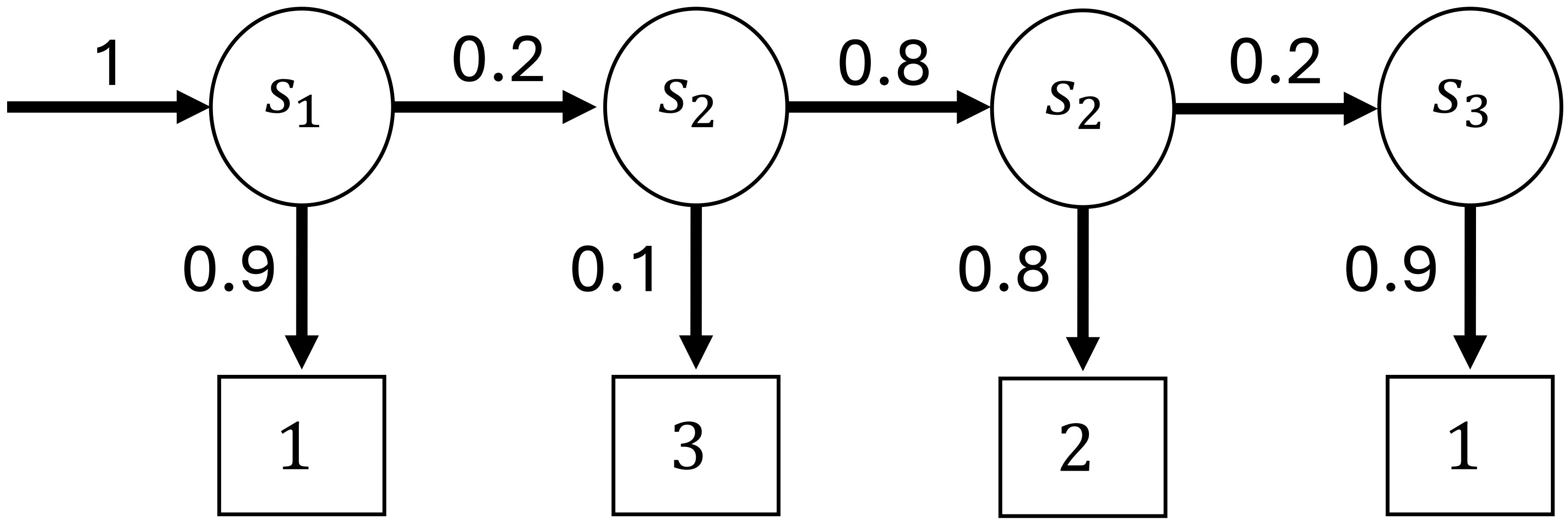


Exercise 2 – Model A

Given:

- $\vec{s}_1 = (s_1, s_2, s_2, s_3)$
- $\vec{o} = (1,3,2,1)$

Probability of $\vec{o}$ given $\vec{s}_1$:
1 * 0.9 * 0.2 * 0.1 * 0.8 * 0.8 * 0.2 * 0.9
= 0.0020736

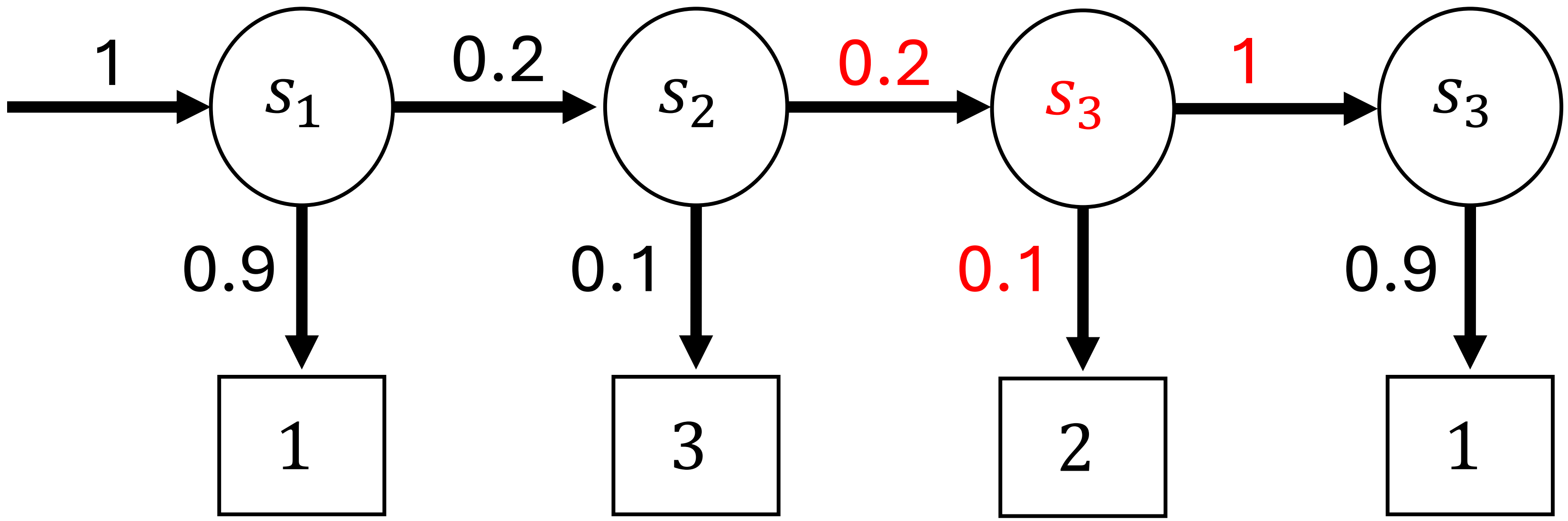


Exercise 2 – Model A

Given:

- $\vec{s}_2 = (s_1, s_2, s_3, s_3)$
- $\vec{o} = (1,3,2,1)$

Probability of $\vec{o}$ given $\vec{s}_2$:
1 * 0.9 * 0.2 * 0.1 * 0.2 * 0.1 * 1 * 0.9
= 0.000324



Exercise 2 - Likelihood

- Total likelihood of obserserving $\vec{o}$ under model A:

0.0020736 + 0.000324 = 0.0023976

- Analogous for model B:

0.0027648 + 0.006912 = 0.0096768

=> The oberservation is more likely under model B