THE UNIVERSITY OF MELBOURNE

Comp90042 Workshop Week 4

23 March

Table of Content

Table of Content

1. Part-of Speech Tagging
2. Hidden Markov Model

Definition

What is **Part-of-Speech (POS) tag**?

Definition

What is **Part-of-Speech (POS) tag**?

- Label of word’s grammatical (primarily syntactic) properties of in the sentence.

POS & Lemmatisation

```
lemmatizer.lemmatize('feeling','v')
```

POS & Lemmatisation

```
lemmatizer.lemmatize('feeling','v')
```

```
'feel'
```

POS & Lemmatisation

```
lemmatizer.lemmatize('feeling','v')
```

```
'feel'
```

```
lemmatizer.lemmatize('feeling','n')
```

POS & Lemmatisation

```
lemmatizer.lemmatize('feeling','v')
```

```
'feel'
```

```
lemmatizer.lemmatize('feeling','n')
```

```
'feeling'
```

POS Tag Set

1. Penn Treebank
2. Brown
3. CLAWS/BNC
4. “Universal”

POS Tag Set

Penn Treebank Tag Set

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

POS Tagging

POS tag (by hand) the following sentence:

Pierre Vinken , 61 years old , will join the board as a nonexecutive director Nov. 29 .

POS Tagging

Pierre ?
Vinken
,
61
years
old
,
Will
join
the
board
as
a
nonexecutive
director
Nov.
29
.

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

Pierre NNP
Vinken
,
61
years
old
,
Will
join
the
board
as
a
nonexecutive
director
Nov.
29
.

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

Pierre	NNP
Vinken	?
,	
61	
years	
old	
,	
Will	
join	
the	
board	
as	
a	
nonexecutive	
director	
Nov.	
29	
.	

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

Pierre	NNP
Vinken	NNP
,	
61	
years	
old	
,	
Will	
join	
the	
board	
as	
a	
nonexecutive	
director	
Nov.	
29	
.	

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

Pierre	NNP
Vinken	NNP
,	?
61	
years	
old	
,	
Will	
join	
the	
board	
as	
a	
nonexecutive	
director	
Nov.	
29	
.	

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

Pierre	NNP
Vinken	NNP
,	,
61	
years	
old	
,	
Will	
join	
the	
board	
as	
a	
nonexecutive	
director	
Nov.	
29	
.	

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

Pierre	NNP
Vinken	NNP
,	,
61	?
years	
old	
,	
Will	
join	
the	
board	
as	
a	
nonexecutive	
director	
Nov.	
29	
.	

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

Pierre	NNP
Vinken	NNP
,	,
61	CD
years	
old	
,	
Will	
join	
the	
board	
as	
a	
nonexecutive	
director	
Nov.	
29	
.	

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

Pierre	NNP
Vinken	NNP
,	,
61	CD
years	?
old	
,	
Will	
join	
the	
board	
as	
a	
nonexecutive	
director	
Nov.	
29	
.	

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

Pierre	NNP
Vinken	NNP
,	,
61	CD
years	NNS
old	
,	
Will	
join	
the	
board	
as	
a	
nonexecutive	
director	
Nov.	
29	
.	

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

Pierre	NNP
Vinken	NNP
,	,
61	CD
years	NNS
old	JJ
,	,
Will	MD
join	VB
the	DT
board	NN
as	IN
a	DT
nonexecutive	JJ
director	NN
Nov.	NNP
29	CD
.	.

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Tag set may vary in different datasets!

POS Tagging

What are some common approaches to POS tagging?

POS Tagging

What are some common approaches to POS tagging?

- Unigram: Use the most common observation on each token.

POS Tagging

What are some common approaches to POS tagging?

- Unigram: Use the most common observation on each token.
  - Corpus: feeling -> nous:14
    verb:25
    adj:5

We always consider feeling is a ?

POS Tagging

What are some common approaches to POS tagging?

- Unigram: Use the most common observation on each token.
  - Corpus: feeling -> nous:14
verb:25
adj:5

We always consider feeling is a verb

POS Tagging

What are some common approaches to POS tagging?

- Unigram: Use the most common observation on each token.

- N-gram: Use most common tag in the same sequence of n tokens (or tags).

POS Tagging

What are some common approaches to POS tagging?

- Unigram: Use the most common observation on each token.
- N-gram: Use most common tag in the same sequence of n tokens (or tags).
- Rule-based: Write rules that disambiguate unigram tags.
- Sequential: Learn a Hidden Markov Model in a tagged corpus.
- Classifier: Treat as a supervised machine learning problem.

Table of Content

Hidden Markov Model

What are the assumptions in Hidden Markov Model (HMM)?

Hidden Markov Model

What are the assumptions in Hidden Markov Model (HMM)?

1. **Markov assumption**

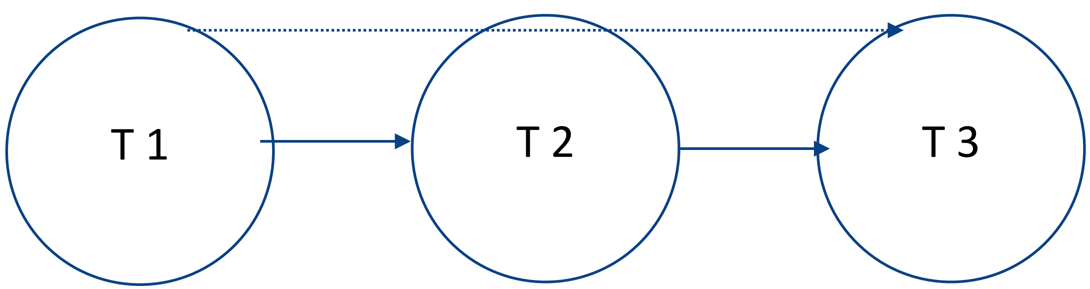
   the likelihood of transitioning into a given state depends only on the current state, and not the previous state(s) (or output(s))

Hidden Markov Model

What are the assumptions in Hidden Markov Model (HMM)?

1. **Markov assumption**
   the likelihood of transitioning into a given state depends only on the current state, and not the previous state(s) (or output(s))

State 3 only depends on state 2, not sate 1



Hidden Markov Model

What are the assumptions in Hidden Markov Model (HMM)?

2. Output independence assumption

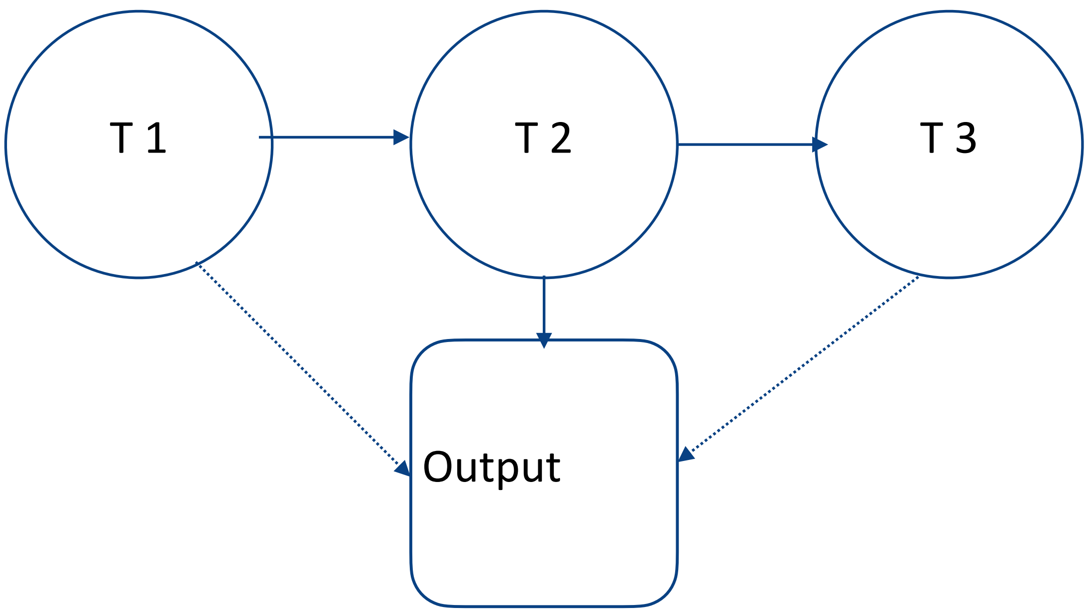
the likelihood of a state producing a certain word (as output) does not depend on the preceding (or following) state(s) (or output(s)).

Hidden Markov Model

What are the assumptions in Hidden Markov Model (HMM)?

2. Output independence assumption

the likelihood of a state producing a certain word (as output) does not depend on the preceding (or following) state(s) (or output(s)).



Hidden Markov Model

What are the assumptions in Hidden Markov Model (HMM)?

1. **Markov assumption**

   the likelihood of transitioning into a given state depends only on the current state, and not the previous state(s) (or output(s))

2. **Output independence assumption**

   the likelihood of a state producing a certain word (as output) does not depend on the preceding (or following) state(s) (or output(s)).

Hidden Markov Training

What are the parameters do we need to learn when training HMM?

1. Initial state probabilities π
   record the distribution of tags for the first token of each sentence

2. Transition probabilities A
   record the distribution of tags of the immediately following token

3. Emission probabilities B
   record the distribution of corresponding tokens

Hidden Markov Training

Estimate π, A and B for POS tagging, based on the following corpus:

1. silver-JJ wheels-NNS turn-VBP
2. wheels-NNS turn-VBP right-JJ
3. right-JJ wheels-NNS turn-VBP

Initial state probabilities π

record the distribution of tags of the first token of each sentence

$$\pi[JJ,\ NNS,\ VBP] = [\frac{2}{3},\ \frac{1}{3},\ 0]$$

Hidden Markov Training

Estimate π, A and B for POS tagging, based on the following corpus:

1. silver-JJ wheels-NNS turn-VBP
2. wheels-NNS turn-VBP right-JJ
3. right-JJ wheels-NNS turn-VBP

Transition probabilities A

record the distribution of tags of the immediately following token

	JJ	NNS	VBP
(from) JJ	0	1	0
NNS	0	0	1
VBP	1	0	0

Hidden Markov Training

Estimate π, A and B for POS tagging, based on the following corpus:

1. silver-JJ wheels-NNS turn-VBP
2. wheels-NNS turn-VBP right-JJ
3. right-JJ wheels-NNS turn-VBP

Emission probabilities B

record the distribution of corresponding token

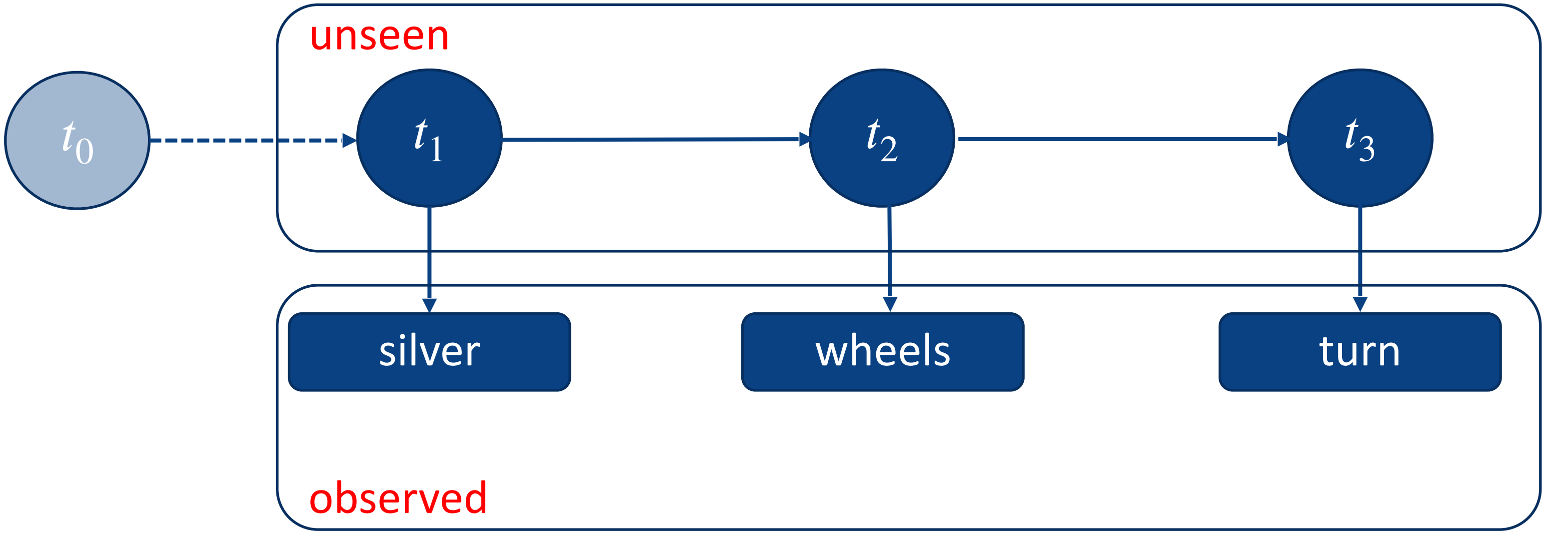
	right	silver	turn	wheels
(from) JJ	$\frac{2}{3}$	$\frac{1}{3}$	0	0
NNS	0	0	0	1
VBP	0	0	1	0

Hidden Markov Inference

Use the following Hidden Markov Model to tag the sentence

silver wheels turn

Visualize the HMM



Hidden Markov Inference

What is **Viterbi algorithm** ?

- a **dynamic programming algorithm** for finding the most likely sequence of hidden states

Use the **Viterbi algorithm** with given parameters below to find the optimal tag sequence.

$$\pi[\mathrm{JJ}, \mathrm{NNS}, \mathrm{VBP}] = [0.3, 0.4, 0.3]$$

A	JJ	NNS	VBP	B	silver	wheels	turn
JJ	0.4	0.5	0.1	JJ	0.8	0.1	0.1
NNS	0.1	0.4	0.5	NNS	0.3	0.4	0.3
VBP	0.4	0.5	0.1	VBP	0.1	0.3	0.6

$\pi[\text{JJ}, \text{NNS}, \text{VBP}] = [0.3, 0.4, 0.3]$

A	JJ	NNS	VBP
JJ	0.4	0.5	0.1
NNS	0.1	0.4	0.5
VBP	0.4	0.5	0.1

B	silver	wheels	turn
JJ	0.8	0.1	0.1
NNS	0.3	0.4	0.3
VBP	0.1	0.3	0.6

Position 1: $\pi \times B(w_1)$

α	1:*silver*		2:*wheels*	3:*turn*
JJ:	JJ	$\pi[\text{JJ}]B[\text{JJ}, \text{silver}]$		
NNS:	NNS	$\pi[\text{NNS}]B[\text{NNS}, \text{silver}]$		
VBP:	VBP	$\pi[\text{VBP}]B[\text{VBP}, \text{silver}]$		

$$\pi[\text{JJ}, \text{NNS}, \text{VBP}] = [0.3, 0.4, 0.3]$$

A	JJ	NNS	VBP	B	silver	wheels	turn
JJ	0.4	0.5	0.1	JJ	0.8	0.1	0.1
NNS	0.1	0.4	0.5	NNS	0.3	0.4	0.3
VBP	0.4	0.5	0.1	VBP	0.1	0.3	0.6

Position i $(i > 1)$:

1. At each cell (t_{i-1}), for every possible tag t_i, compute $A[t_{i-1}, t_i] \times B[t_t, w_i]$

2. Choose t_i which produce highest prob. for the assumed t_{i-1}

α	*1:silver*	*2:wheels*		*3:turn*
JJ:	0.24	JJ $\rightarrow$ JJ 0.24	$A[\text{JJ,JJ}]B[\text{JJ, wheels}]$ $\boxed{\times 0.4 \times 0.1 = \mathbf{0.0096}}$	
		NNS $\rightarrow$ JJ 0.12	$A[\text{NNS,JJ}]B[\text{JJ, wheels}]$ $\times 0.1 \times 0.1 = 0.0012$	
		VBP $\rightarrow$ JJ 0.03	$A[\text{VBP,JJ}]B[\text{JJ, wheels}]$ $\times 0.4 \times 0.1 = 0.0012$	
NNS:	0.12	JJ $\rightarrow$ NNS 0.24	$A[\text{JJ,NNS}]B[\text{NNS, wheels}]$ $\boxed{\times 0.5 \times 0.4 = \mathbf{0.048}}$	
		NNS $\rightarrow$ NNS 0.12	$A[\text{NNS,NNS}]B[\text{NNS, wheels}]$ $\times 0.4 \times 0.4 = 0.0192$	
		VBP $\rightarrow$ NNS 0.03	$A[\text{VBP,NNS}]B[\text{NNS, wheels}]$ $\times 0.5 \times 0.4 = 0.006$	
VBP:	0.03	JJ $\rightarrow$ VBP 0.24	$A[\text{JJ,VBP}]B[\text{VBP, wheels}]$ $\times 0.1 \times 0.3 = 0.0072$	
		NNS $\rightarrow$ VBP 0.12	$A[\text{NNS,VBP}]B[\text{VBP, wheels}]$ $\boxed{\times 0.5 \times 0.3 = \mathbf{0.018}}$	
		VBP $\rightarrow$ VBP 0.03	$A[\text{VBP,VBP}]B[\text{VBP, wheels}]$ $\times 0.1 \times 0.3 = 0.0009$	

highest prob.

Repeat previous step until reach end of sentence

α	*1:silver*	*2:wheels*		*3:turn*
JJ:	0.24	0.0096 JJ $\rightarrow$ JJ	JJ $\rightarrow$ JJ 0.0096 NNS $\rightarrow$ JJ 0.048 VBP $\rightarrow$ JJ 0.018	A[JJ,JJ]B[JJ, turn] $\times 0.4 \times 0.1 = 0.000384$ A[NNS,JJ]B[JJ, turn] $\times 0.1 \times 0.1 = 0.00048$ A[VBP,JJ]B[JJ, turn] $\times 0.4 \times 0.1 = \mathbf{0.00072}$
NNS:	0.12	0.048 JJ $\rightarrow$ NNS	JJ $\rightarrow$ NNS 0.0096 NNS $\rightarrow$ NNS 0.048 VBP $\rightarrow$ NNS 0.018	A[JJ,NNS]B[NNS, turn] $\times 0.5 \times 0.3 = 0.00144$ A[NNS,NNS]B[NNS, turn] $\times 0.4 \times 0.3 = \mathbf{0.00576}$ A[VBP,NNS]B[NNS, turn] $\times 0.5 \times 0.3 = 0.0027$
VBP:	0.03	0.018 NNS $\rightarrow$ VBP	JJ $\rightarrow$ VBP 0.0096 NNS $\rightarrow$ VBP 0.048 VBP $\rightarrow$ VBP 0.018	A[JJ,VBP]B[VBP, turn] $\times 0.1 \times 0.6 = 0.000576$ A[NNS,VBP]B[VBP, turn] $\times 0.5 \times 0.6 = \mathbf{0.0144}$ A[VBP,VBP]B[VBP, turn] $\times 0.1 \times 0.6 = 0.00108$

Only optimal sub-sequences are stored, with the back pointer to previous tag of this cell

Hidden Markov Inference

What is the **time complexity** of Viterbi algorithm ?

$$O(T^2 W)$$

For each word $w_i \in W$, T possible previous tags and T current tags are considered.

Tags: 100
Tokens: 10
Time complexity: ~100,000
It's practical

Takeaways

1. Part-of-Speech tagging
   - Definition (including common POS tag labels)
   - Approaches

2. Hidden Markov Model (HMM)
   - Assumptions
   - Training (MLE)
   - Inference (Viterbi)

THE UNIVERSITY OF MELBOURNE

Thank you