

Language Modeling

Applications of Language Modeling



What is the|



what is the **fastest car in the world**

what is the **meaning of life**

what is the **keto diet**

what is the **quadratic formula**

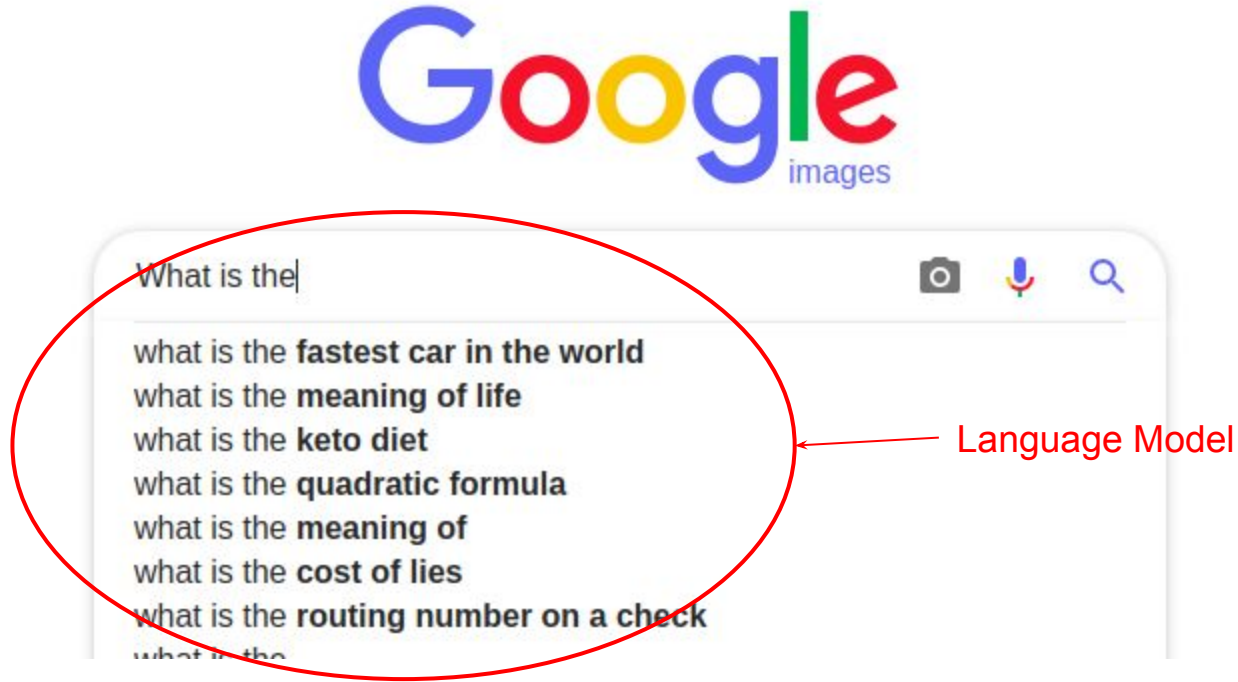
what is the **meaning of**

what is the **cost of lies**

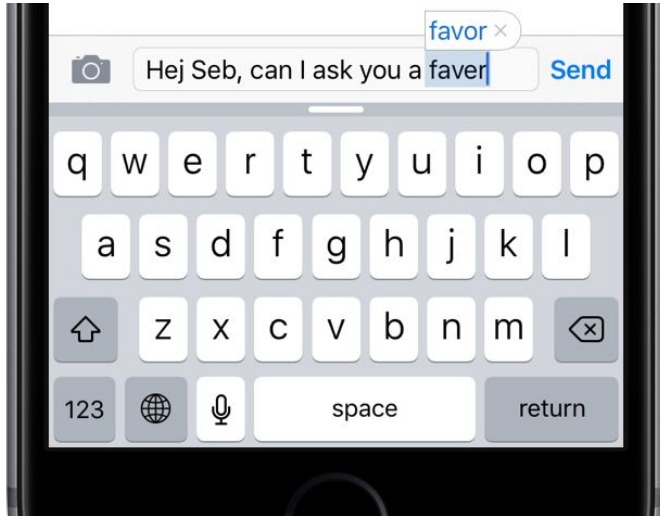
what is the **routing number on a check**

what is the

Applications of Language Modeling

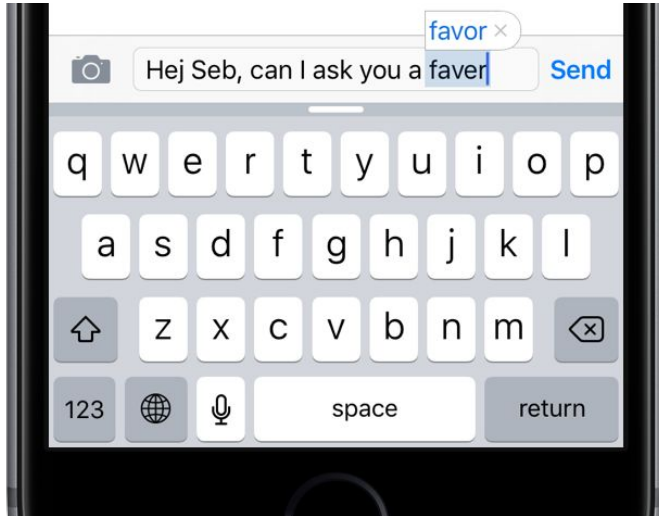


Applications of Language Modeling

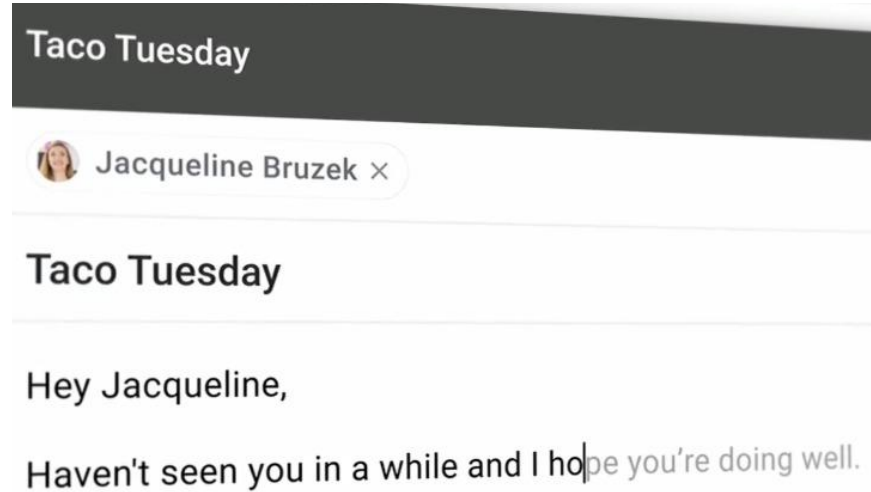


Auto-Correct

Applications of Language Modeling



Auto-Correct



Email Completion

Applications of Language Modeling

Handwriting
Recognition

my	alarm	clock	did	not
my	alarm	code	soil	rout
		circle	raid	hot
		shute	risk	riot
		clock	visit	not
			did	must

wake	me	up	this	morning
wake	me	up	thai	moving
			taxis	having
			this	running
			tier	morning
				loving

Applications of Language Modeling

Handwriting Recognition

my alarm clock did not
my alarm code circle soil rout
shute risk hot
clock visit riot
did must

wake me up this morning
wake me up thai taxis moving
this having
tier running
loving

Machine Translation

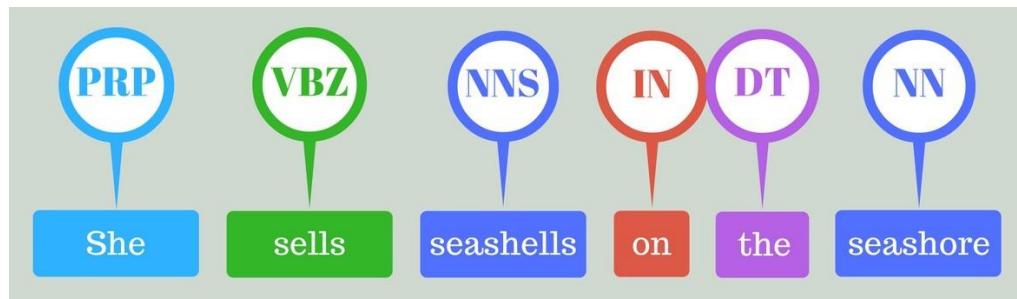


Applications of Language Modeling

Handwriting
Recognition

my alarm clock did not
my alarm code circle shute soil raid rout
clock risk visit hot
did must
wake me up this morning
wake me up thai taxi moving
this tier having
morning running
loving

Machine
Translation



POS Tags

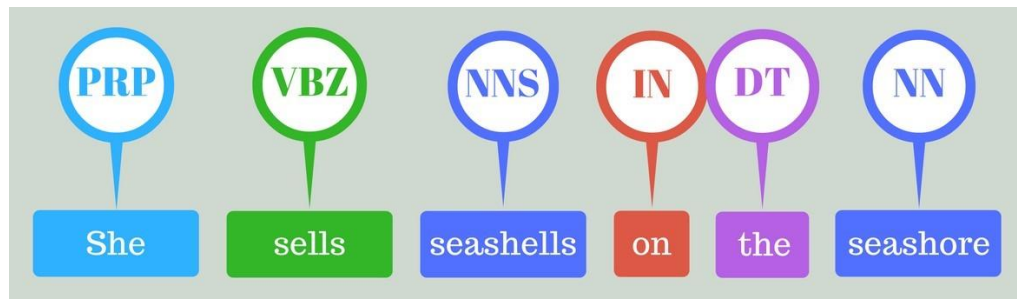
Applications of Language Modeling

Handwriting
Recognition

my alarm clock did not
my alarm code circle soil rout
shute risk riot
clock visit did not
must

wake me up this morning
wake me up thai moving
taxi having
this running
tier morning
loving

Machine
Translation



POS Tags



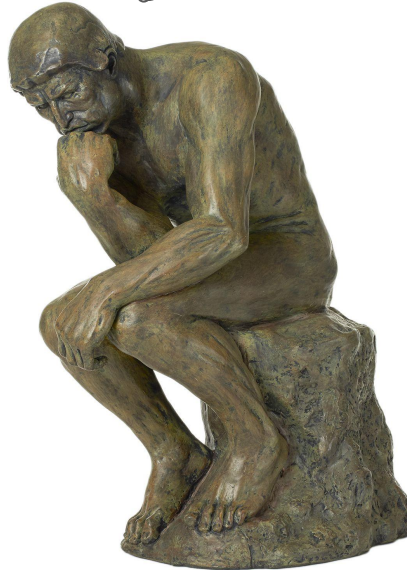
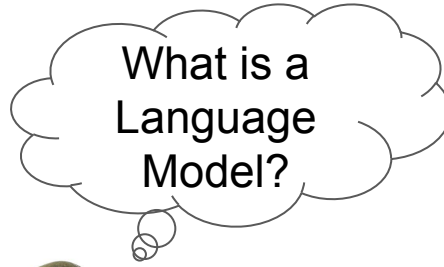
Speech
Recognition

Module Topics

- What is a Language Model in NLP?
- N-gram Language Model
- Implementing an N-gram Language Model
- Neural Language Model
- Implementing a Neural Language Model

What is a Language Model in NLP?

What is a Language Model in NLP?



What is a Language Model in NLP?

A language model learns to predict the probability of a sequence of words.

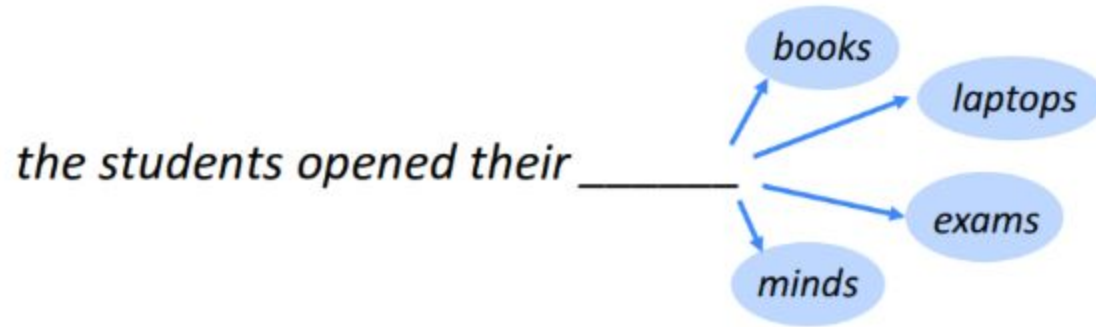
What is a Language Model in NLP?

A language model learns to predict the probability of a sequence of words.

the students opened their _____

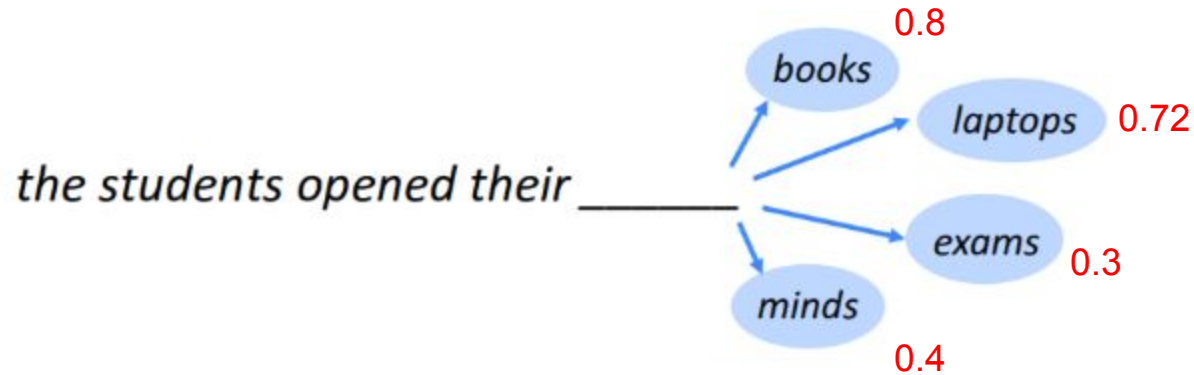
What is a Language Model in NLP?

A language model learns to predict the probability of a sequence of words.



What is a Language Model in NLP?

A language model learns to predict the probability of a sequence of words.



What is a Language Model in NLP?

A language model learns to predict the probability of a sequence of words.

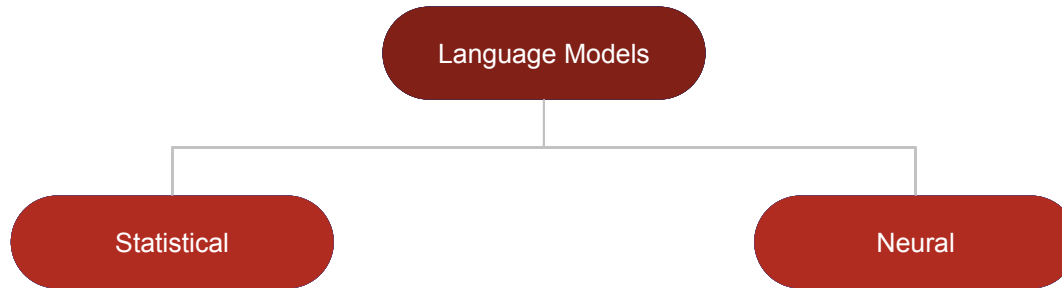
Why?

- $P(\text{I saw a van}) > P(\text{eyes awe of an}) \rightarrow \text{Speech Recognition}$
- $P(\text{high winds tonite}) > P(\text{large winds tonite}) \rightarrow \text{Machine Translation}$
- $P(\text{about fifteen minutes from}) > P(\text{about fifteen minuets from}) \rightarrow \text{Spell Correction}$
- And many more..

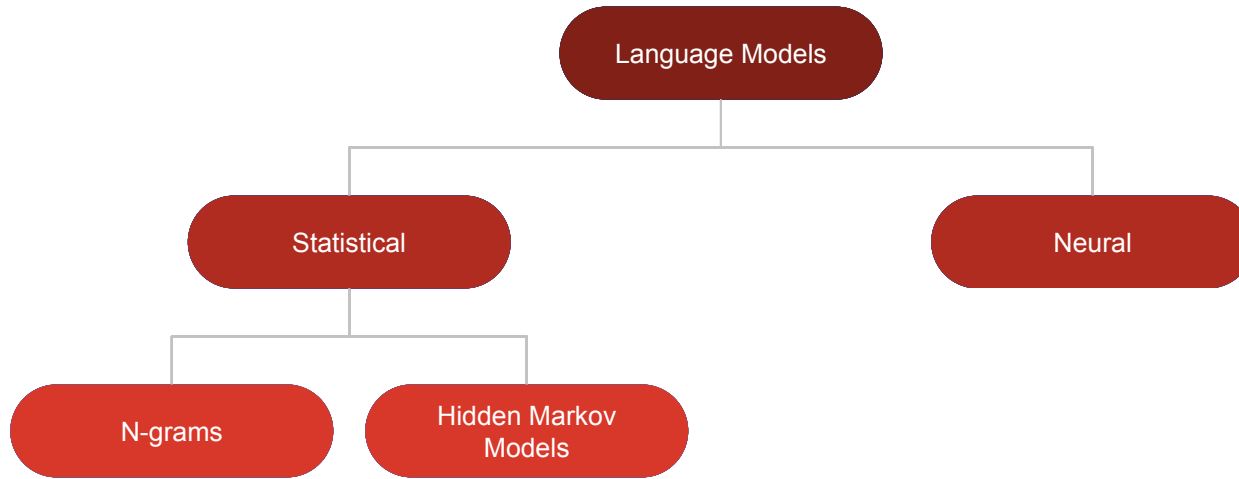
What is a Language Model in NLP?

Language Models

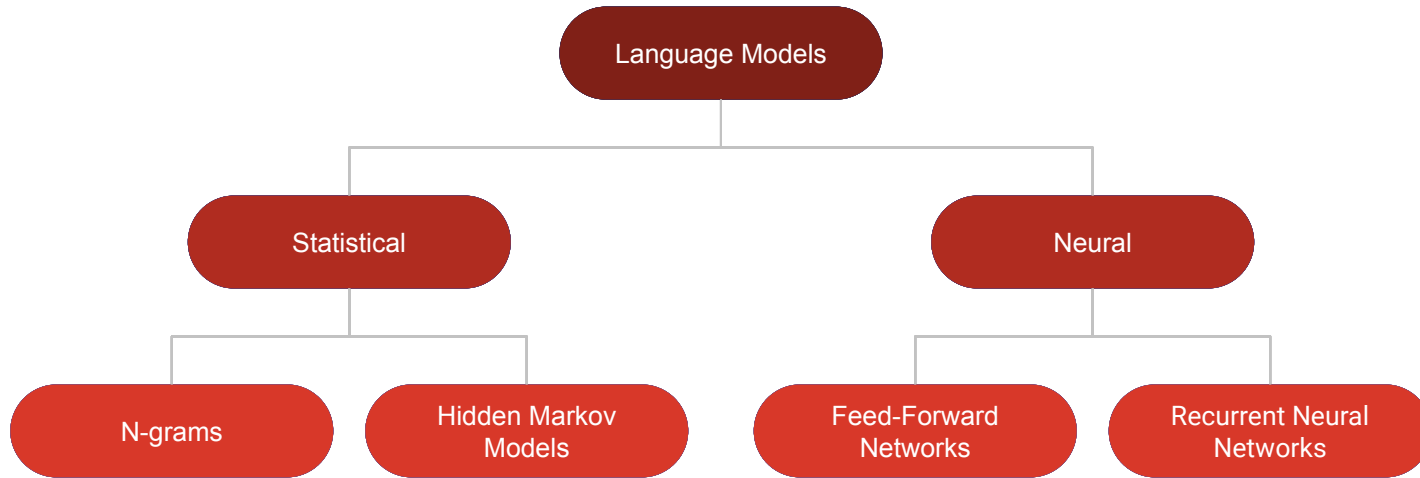
What is a Language Model in NLP?



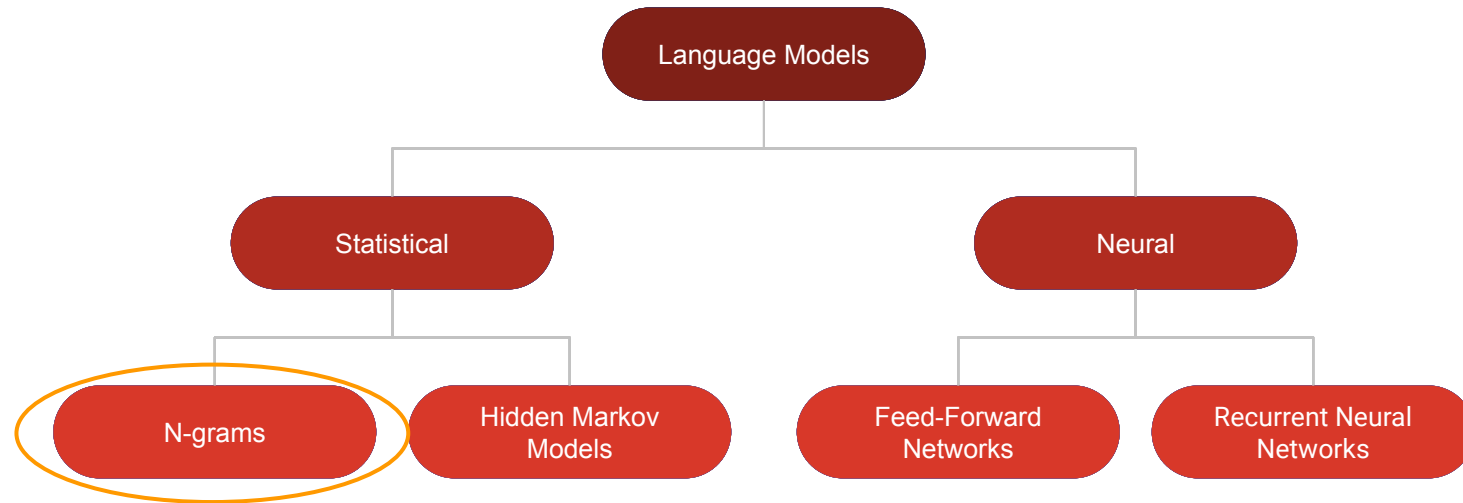
What is a Language Model in NLP?



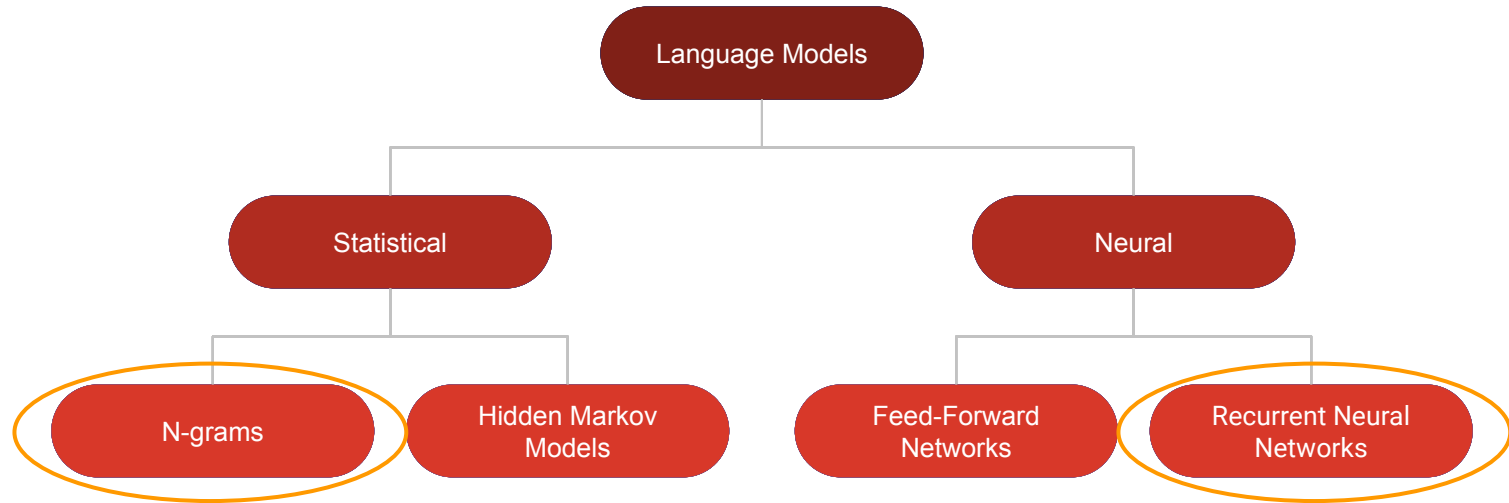
What is a Language Model in NLP?



What is a Language Model in NLP?



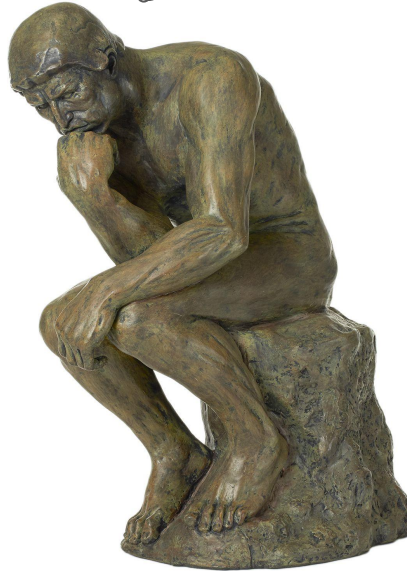
What is a Language Model in NLP?



N-gram Language Model

N-gram Language Model

What is
N-gram?



N-gram Language Model

I love reading blogs about data science on Analytics Vidhya.

N-gram Language Model

I love reading blogs about data science on Analytics Vidhya.

- An N-gram is a sequence of N tokens (or words).

N-gram Language Model

I love reading blogs about data science on Analytics Vidhya.

- An N-gram is a sequence of N tokens (or words).

N=1 [I, love, reading, blogs, about, data, science, on, Analytics, Vidhya]

N-gram Language Model

I love reading blogs about data science on Analytics Vidhya.

- An N-gram is a sequence of N tokens (or words).

N=1 [I, love, reading, blogs, about, data, science, on, Analytics, Vidhya]

N=2 [I love, love reading, reading blogs, blogs about...]

N-gram Language Model

I love reading blogs about data science on Analytics Vidhya.

- An N-gram is a sequence of N tokens (or words).

N=1 [I, love, reading, blogs, about, data, science, on, Analytics, Vidhya]

N=2 [I love, love reading, reading blogs, blogs about...]

N=3 [I love reading, love reading blogs, reading blogs about...]

N-gram Language Model

I love reading blogs about data science on Analytics Vidhya.

- An N-gram is a sequence of N tokens (or words).

N=1	[I, love, reading, blogs, about, data, science, on, Analytics, Vidhya]	Unigrams
N=2	[I love, love reading, reading blogs, blogs about...]	Bigrams
N=3	[I love reading, love reading blogs, reading blogs about...]	Trigrams

N-gram Language Model

Can you please come



The diagram illustrates the concept of history in an N-gram language model. A horizontal line with brackets at both ends spans the width of the sentence "Can you please come". An upward-pointing arrow originates from the word "please" and points to the center of this bracketed line, indicating that the current word is being predicted based on the preceding words in the sequence.

N-gram Language Model

Can you please come

History

Word being predicted

N-gram Language Model



N-gram Language Model



N-gram Language Model

We need:

N-gram Language Model

We need:

- $P(\text{here} \mid \text{can, you, please, come})$

N-gram Language Model

We need:

- $P(\text{here} \mid \text{can, you, please, come}) = P(w_n \mid w_1, w_2, \dots, w_{n-1})$

N-gram Language Model

We need:

- $P(\text{here} \mid \text{can, you, please, come}) = P(w_n \mid w_1, w_2, \dots, w_{n-1})$
- $P(w_n \mid w_1, w_2, \dots, w_{n-1}) \approx P(w_n \mid w_{n-1})$ (markov assumption)

N-gram Language Model

$$P(w_n \mid w_1, w_2, \dots, w_{n-1}) \approx P(w_n \mid w_{n-1})$$

N-gram Language Model

$$P(w_n \mid w_1, w_2, \dots, w_{n-1}) \approx P(w_n \mid w_{n-1})$$



$$P(\text{here} \mid \text{can, you, please, come}) \approx P(\text{here} \mid \text{come})$$

N-gram Language Model

$$P(w_n | w_1, w_2, \dots, w_{n-1}) \approx P(w_n | w_{n-1})$$



$$P(\text{here} | \text{can, you, please, come}) \approx P(\text{here} | \text{come})$$



Can you please come **here**

N-gram Language Model

$$P(w_n \mid w_1, w_2, \dots, w_{n-1}) \approx P(w_n \mid w_{n-1}) \quad \text{“unigram”}$$

$$P(w_n \mid w_1, w_2, \dots, w_{n-1}) \approx P(w_n \mid w_{n-1}, w_{n-2}) \quad \text{“bigram”}$$

Implementing an N-gram Language Model

- <JUPYTER NOTEBOOK>

Neural Language Model

**Limitations of N-grams
approach:**

Neural Language Model

Limitations of N-grams approach:

- Requires large amounts of compute



Neural Language Model

Limitations of N-grams approach:

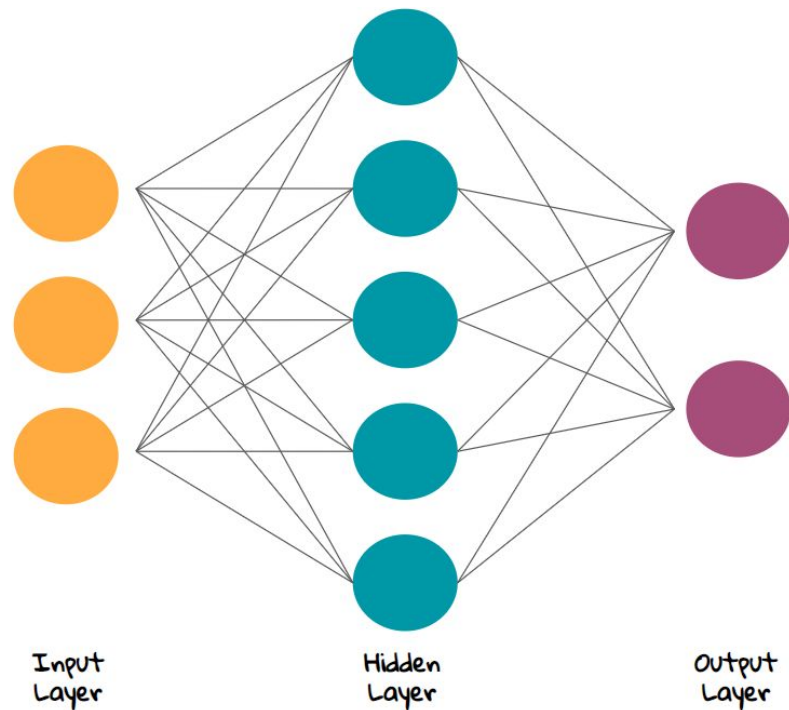
- Requires large amounts of compute
- N-grams are a sparse representation of language.

	i	want	to	eat	chinese	food
i	0.002	0.33	0	0.0036	0	0
want	0.0022	0	0.66	0.0011	0.0065	0.0065
to	0.00083	0	0.0017	0.28	0.00083	0
eat	0	0	0.0027	0	0.021	0.0027
chinese	0.0063	0	0	0	0	0.52
food	0.014	0	0.014	0	0.00092	0.0037
lunch	0.0059	0	0	0	0	0.0029
spend	0.0036	0	0.0036	0	0	0

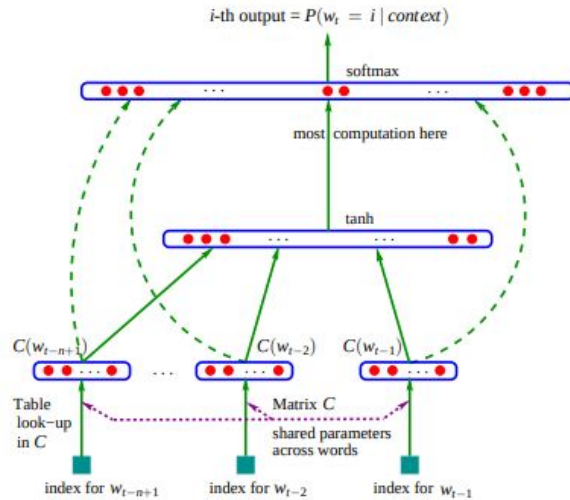
Neural Language Model

Neural Language Modeling:

- Outperformed all previous approaches
- Better ability to generalize
- Multiple approaches:
 - Feed Forward NN
 - CNN
 - RNN etc.

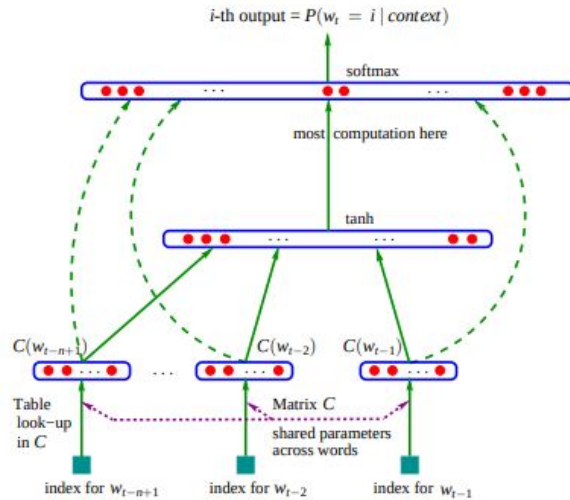


Neural Language Model

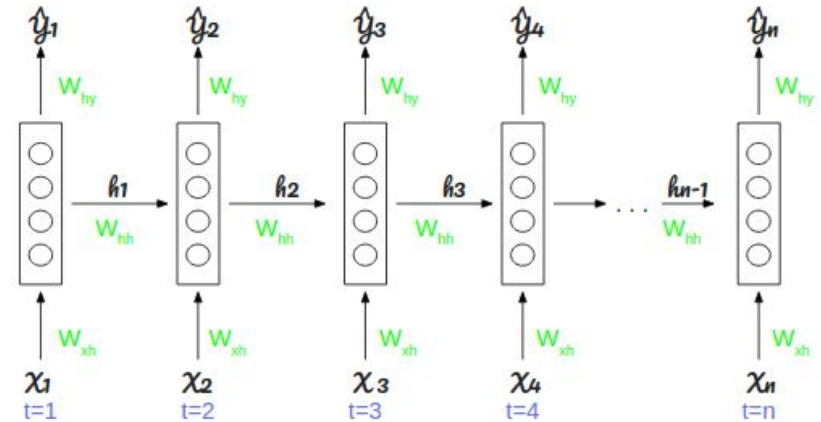


Feed Forward NN

Neural Language Model



Feed Forward NN



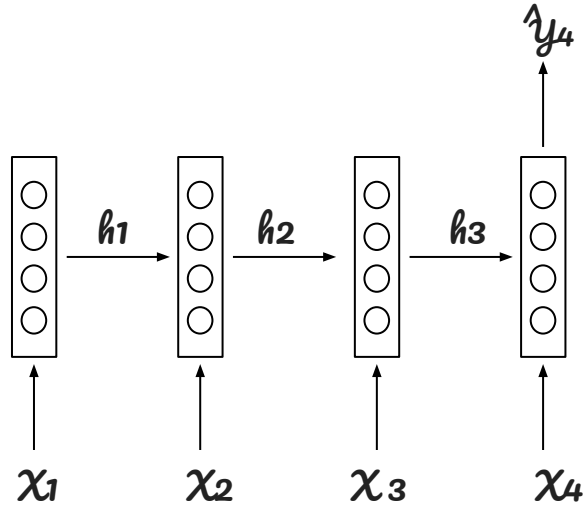
Recurrent NN

Neural Language Model

RNN for Language Modeling:

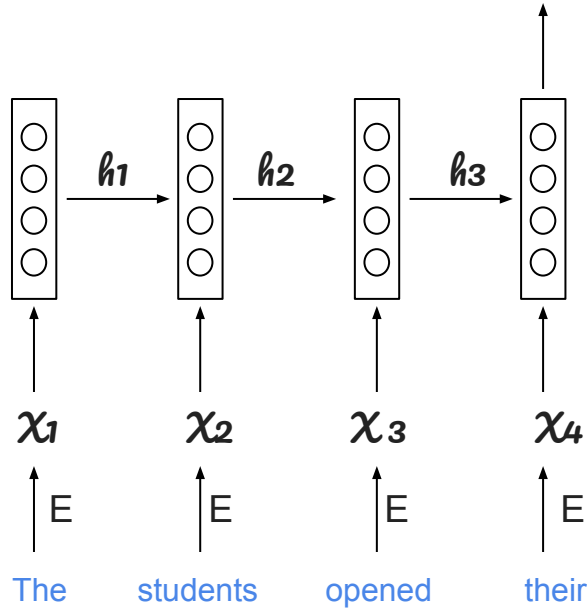
Neural Language Model

RNN for Language Modeling:



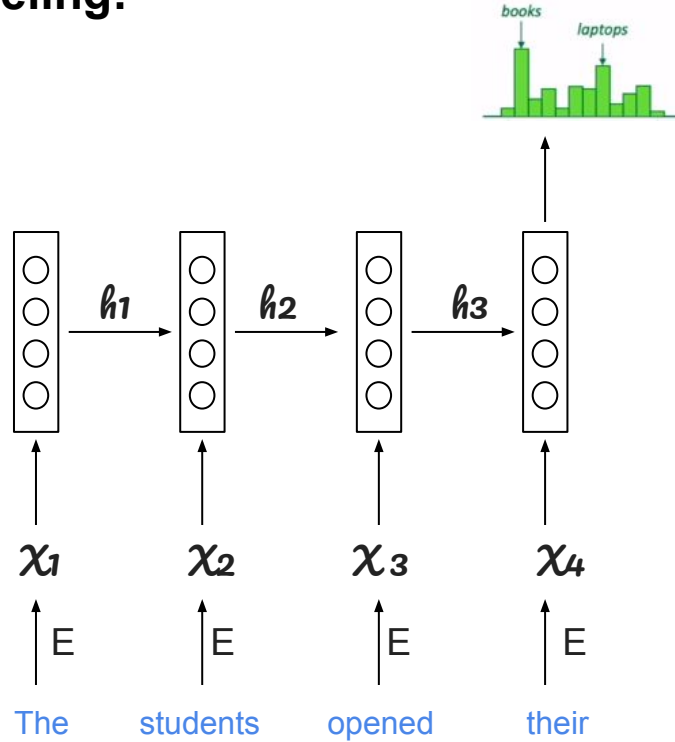
Neural Language Model

RNN for Language Modeling:



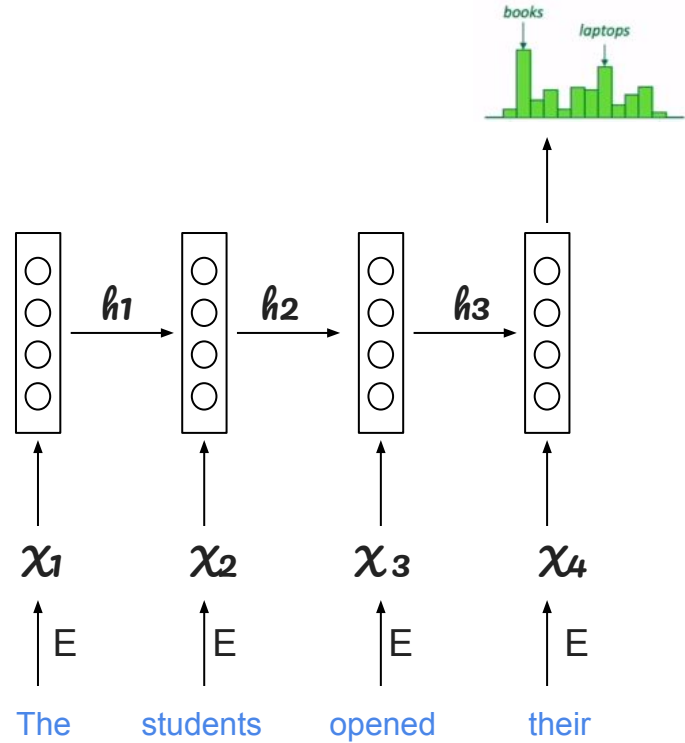
Neural Language Model

RNN for Language Modeling:



Neural Language Model

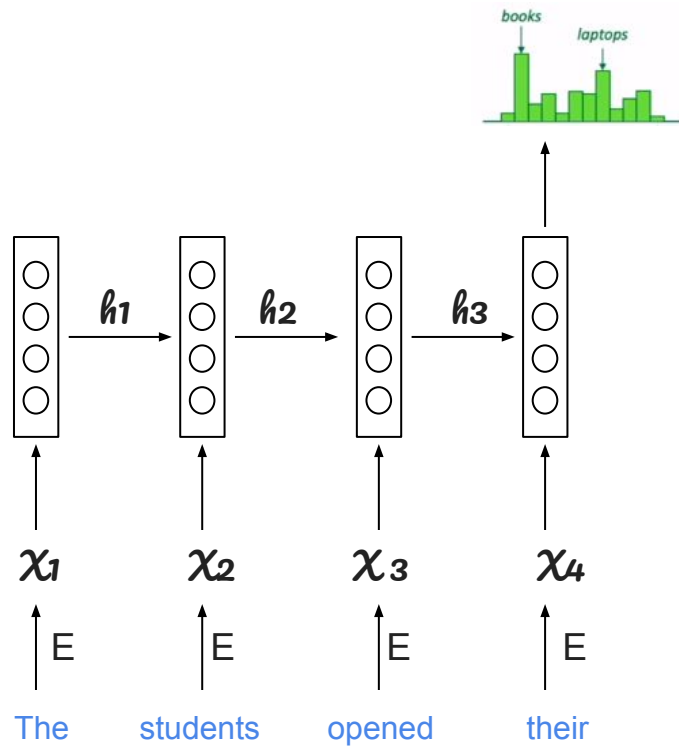
Advantages of RNN for Language Modeling:



Neural Language Model

Advantages of RNN for Language Modeling:

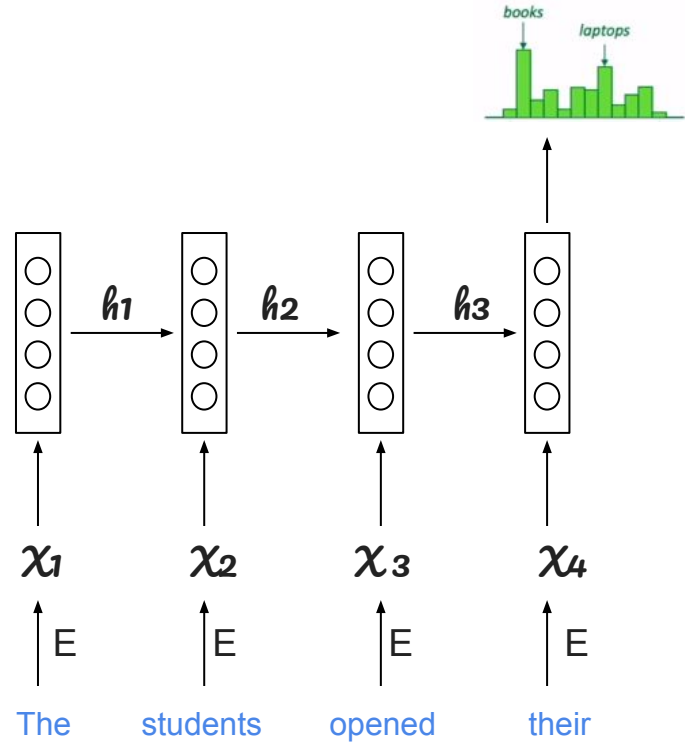
- Variable length input sequences



Neural Language Model

Advantages of RNN for Language Modeling:

- Variable length input sequences
- Combined context
- Fixed computation/model size



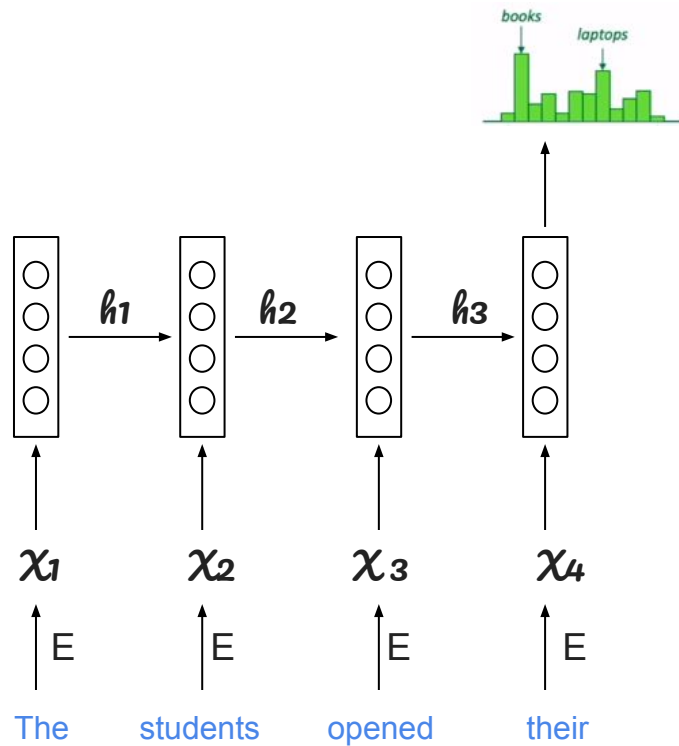
Neural Language Model

Advantages of RNN for Language Modeling:

- Variable length input sequences
- Combined context
- Fixed computation/model size

The cat,..., which already ate,..., was full

Cats,..., which already ate,..., were full



Implementing a Neural Language Model

- <JUPYTER NOTEBOOK>