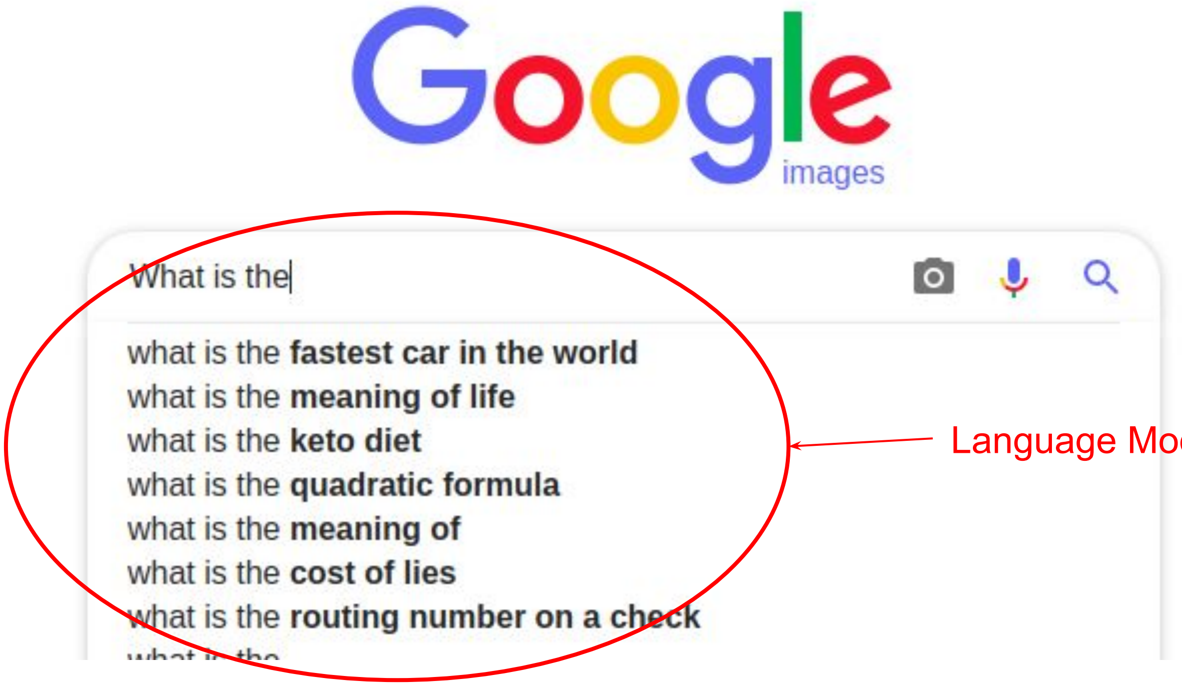
Language Modeling

Analytics Vidhya
Learn everything about analytics

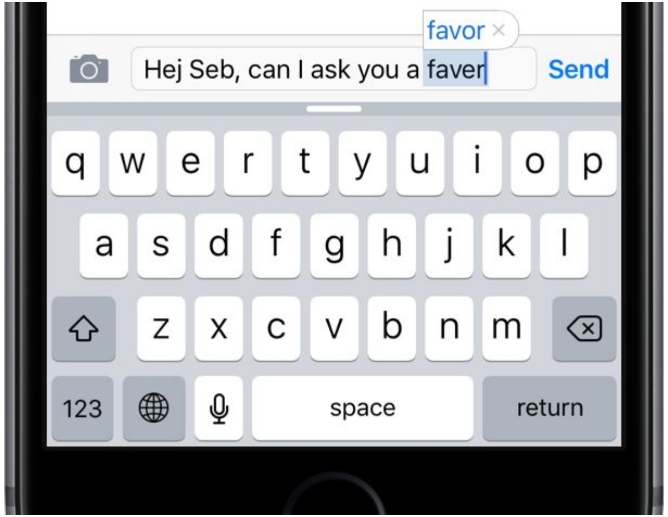
Applications of Language Modeling

Applications of Language Modeling



Language Model

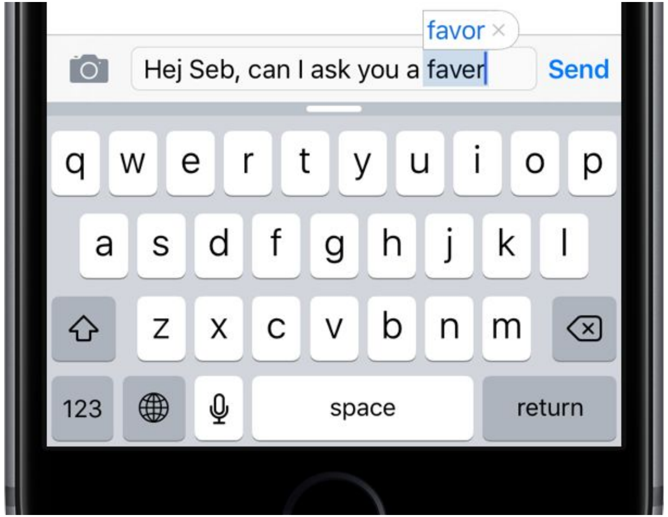
Applications of Language Modeling



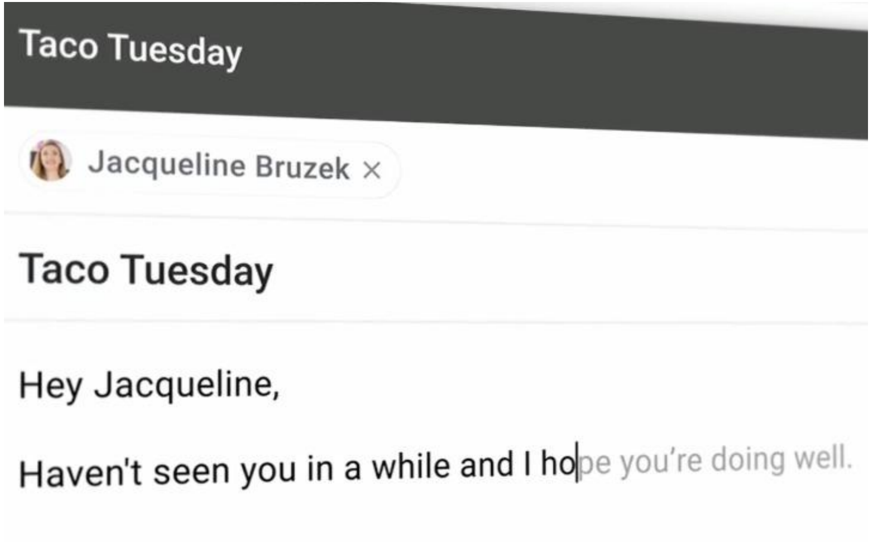
Auto-Correct

Analytics Vidhya
Learn everything about analytics

Applications of Language Modeling



Auto-Correct

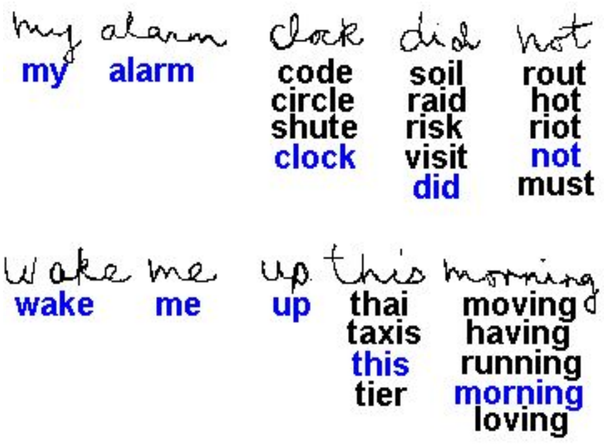


Email Completion

Analytics Vidhya
Learn everything about analytics

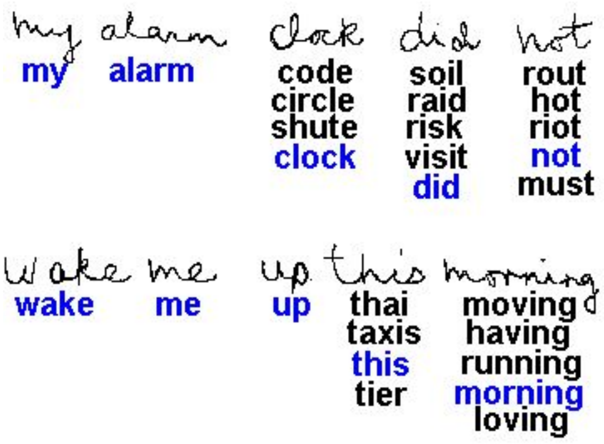
Applications of Language Modeling

Handwriting Recognition



Applications of Language Modeling

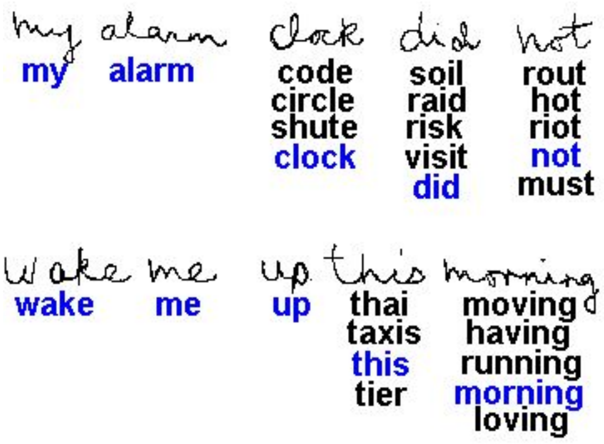
Handwriting Recognition



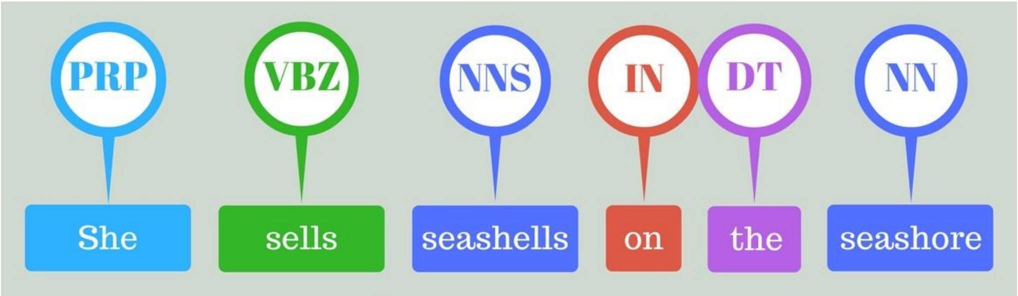
Machine Translation

Applications of Language Modeling

Handwriting Recognition



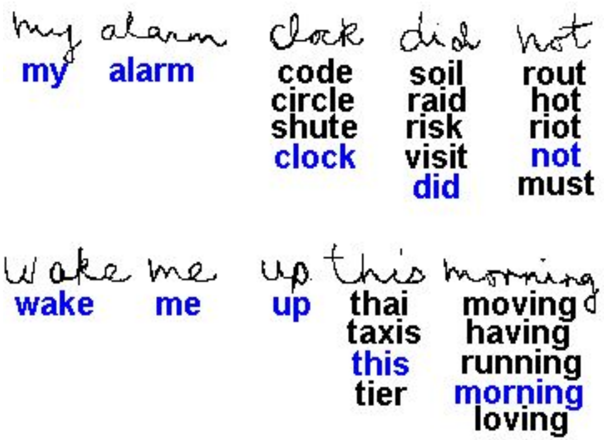
Machine Translation



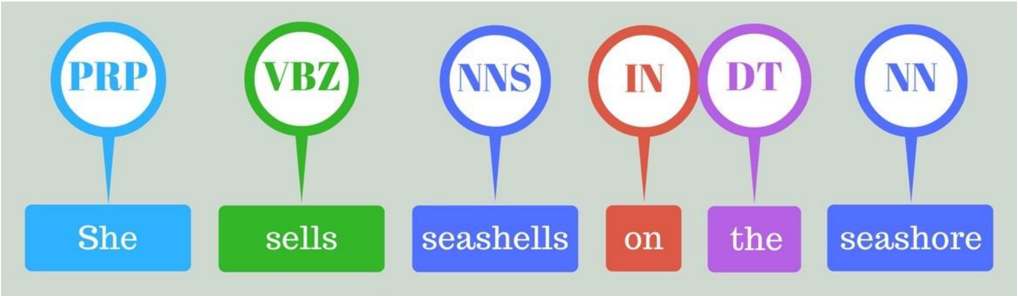
POS Tags

Applications of Language Modeling

Handwriting Recognition



Machine Translation



POS Tags

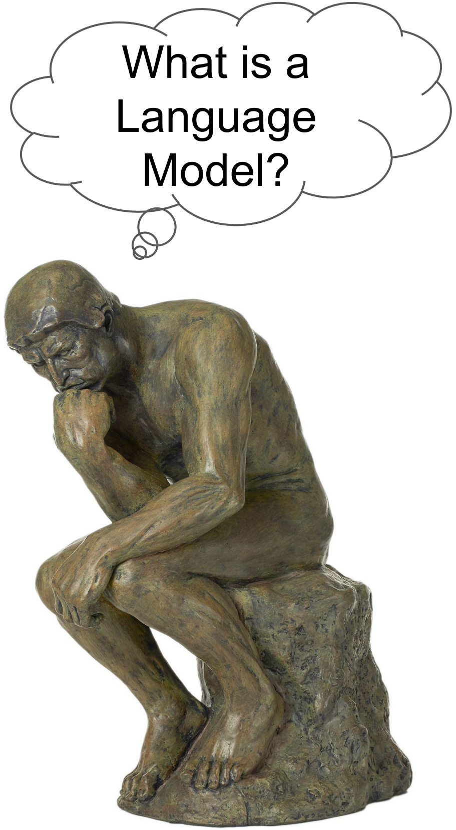
Speech Recognition

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?



Analytics Vidhya
Learn everything about analytics

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?

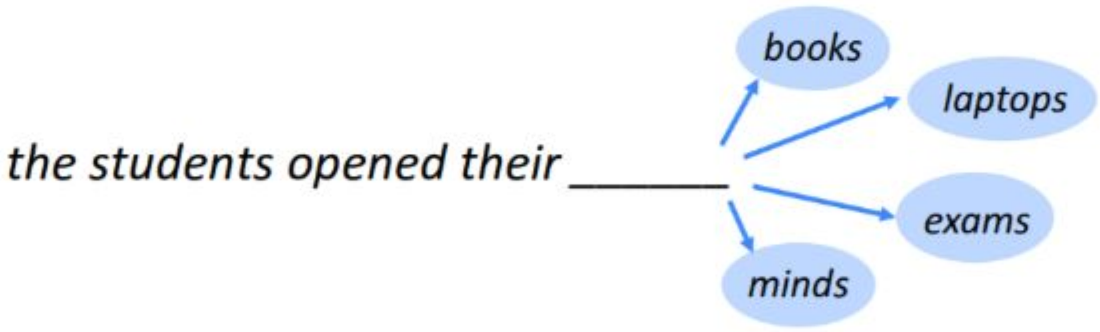
A language model learns to predict the probability of a sequence of words.

the students opened their ______

Analytics Vidhya
Learn everything about analytics

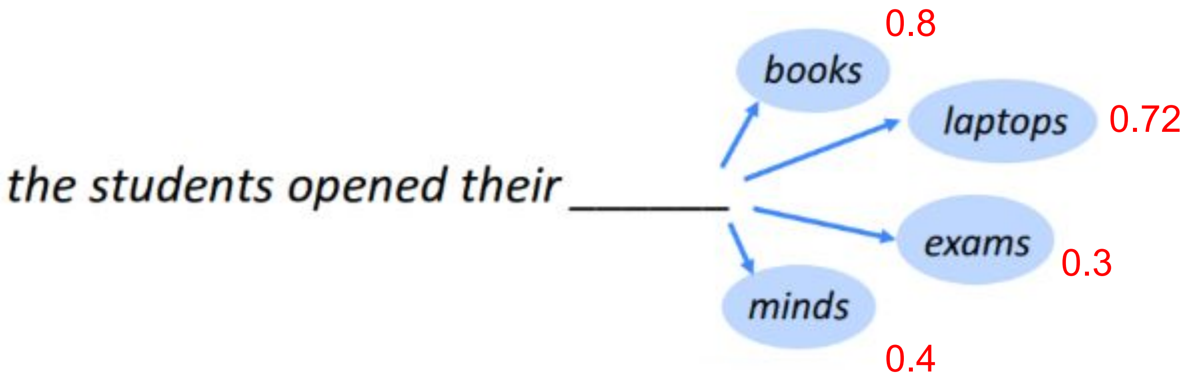
What is a Language Model in NLP?

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?

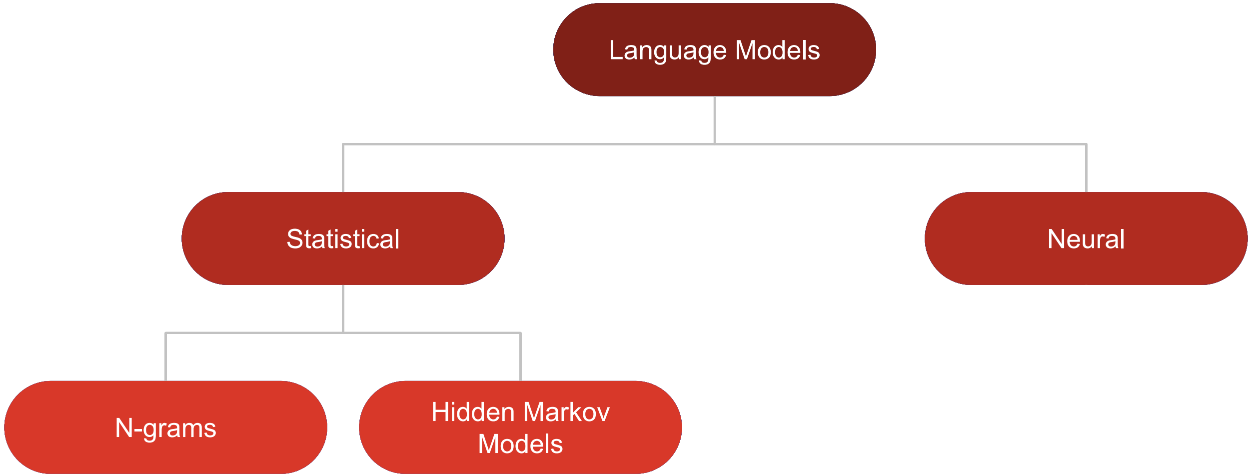
- P(I saw a van) > P(eyes awe of an) → Speech Recognition
- P(**high** winds tonite) > P(**large** winds tonite) → Machine Translation
- P(about fifteen **minutes** from) > P(about fifteen **minuets** from) → Spell Correction
- And many more..

What is a Language Model in NLP?

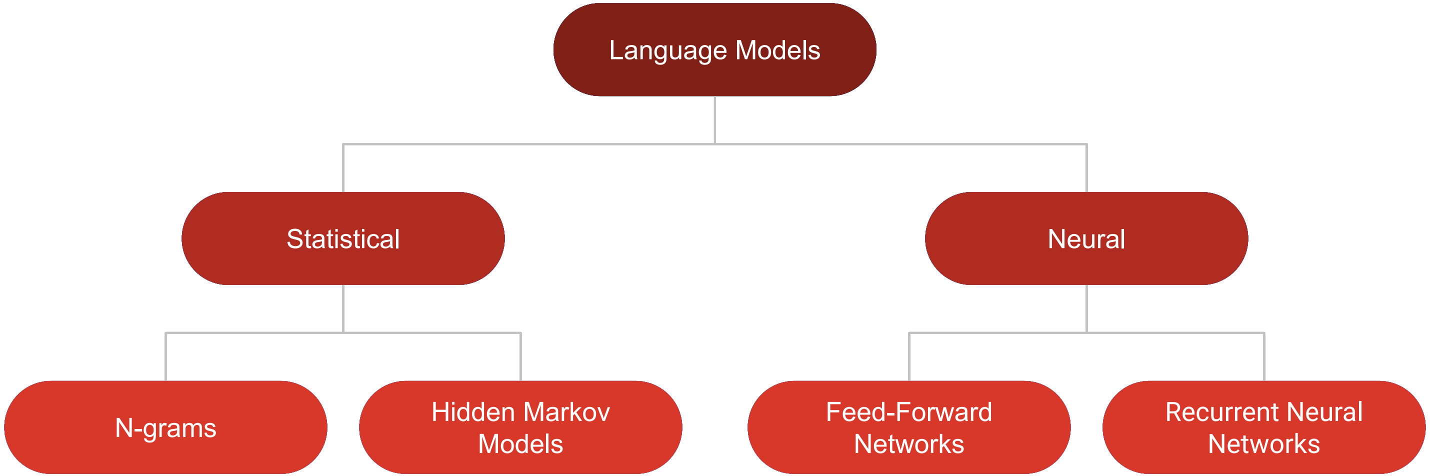
Language Models

What is a Language Model in NLP?

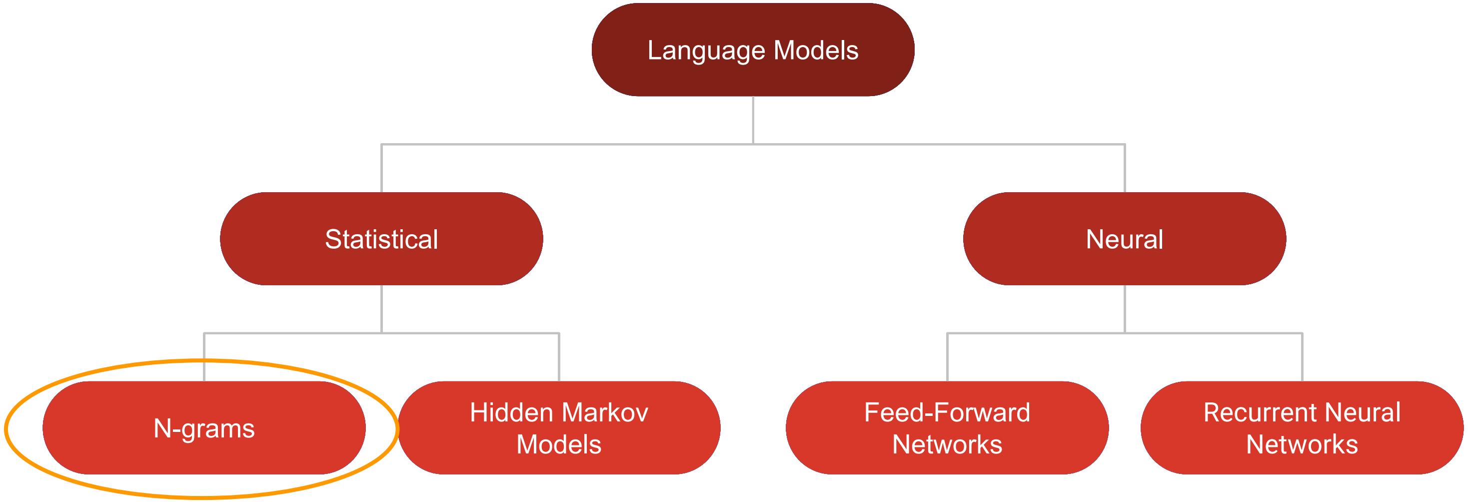
What is a Language Model in NLP?



What is a Language Model in NLP?

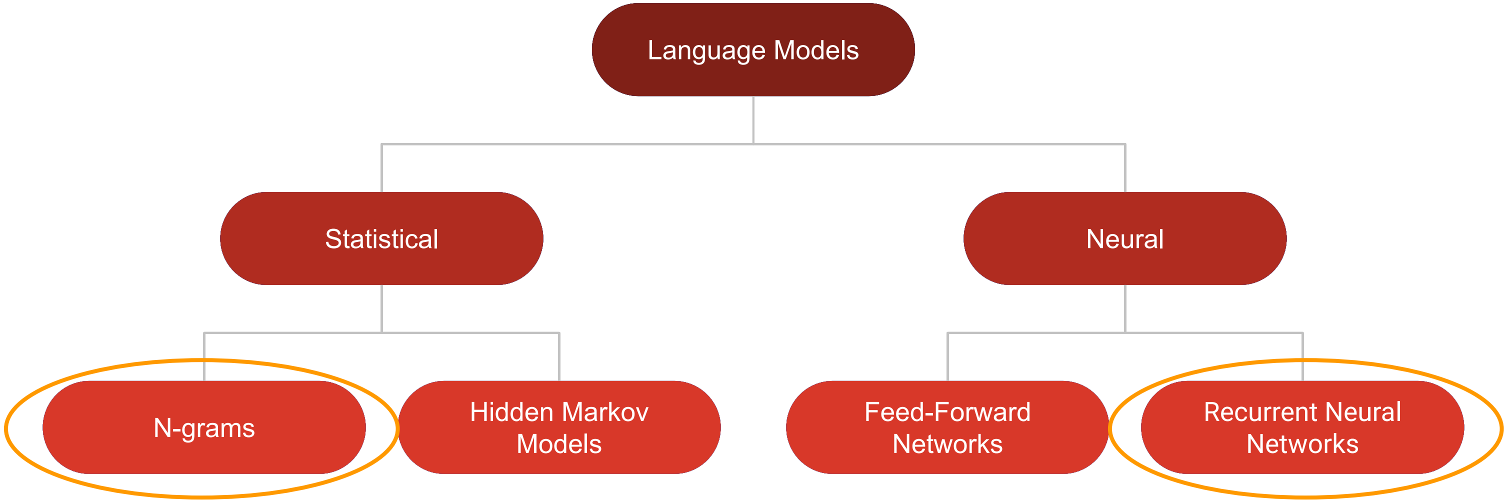


What is a Language Model in NLP?



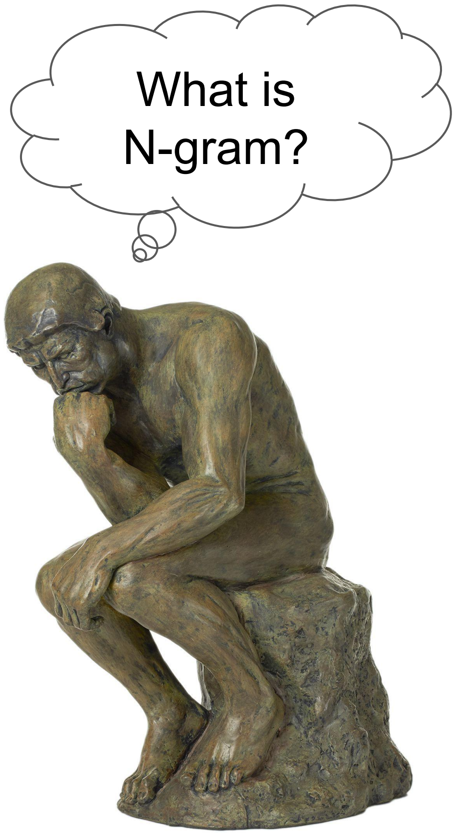
Analytics Vidhya
Learn everything about analytics

What is a Language Model in NLP?



N-gram Language Model

N-gram Language Model



Analytics Vidhya
Learn everything about analytics

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

N-gram Language Model

N-gram Language Model

N-gram Language Model

P(here | can, you , please, come)

$\underbrace{\text{Can you please come}}_{\text{History}}$ here ?

Word being predicted

N-gram Language Model

We need:

N-gram Language Model

We need:

- P(here | can, you , please, come)

N-gram Language Model

We need:

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

N-gram Language Model

We need:

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

N-gram Language Model

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

N-gram Language Model

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

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

N-gram Language Model

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

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

Can you please come here

Analytics Vidhya
Learn everything about analytics

N-gram Language Model

$$P(w_n \mid w_1, w_2, \ldots, w_{n-1}) \approx P(w_n \mid w_{n-1}) \qquad \text{"unigram"}$$

$$P(w_n \mid w_1, w_2, \ldots, w_{n-1}) \approx P(w_n \mid w_{n-1}, w_{n-2}) \qquad \text{"bigram"}$$

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:

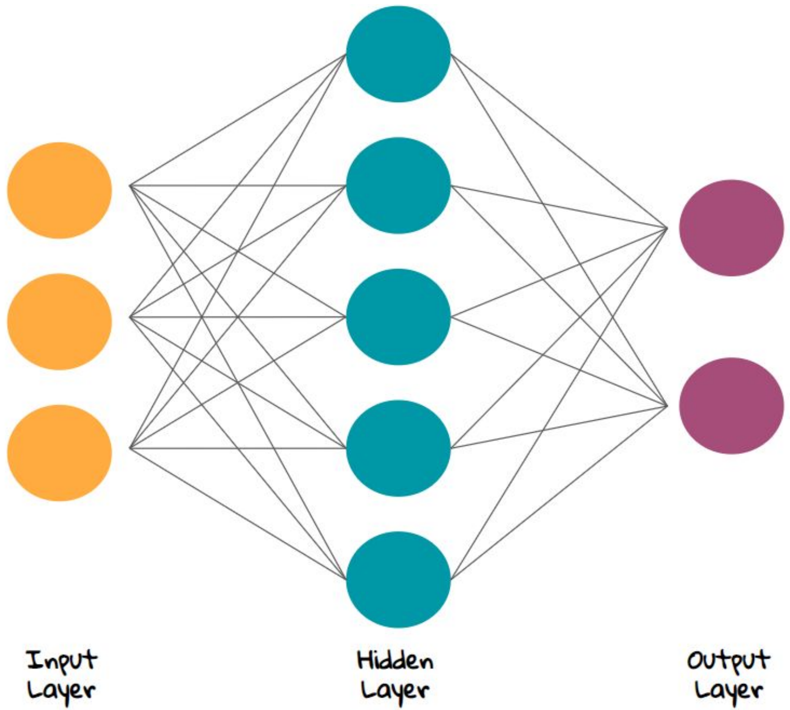
- Requires large amounts of compute
- N-grams are a sparse representation of language.

	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

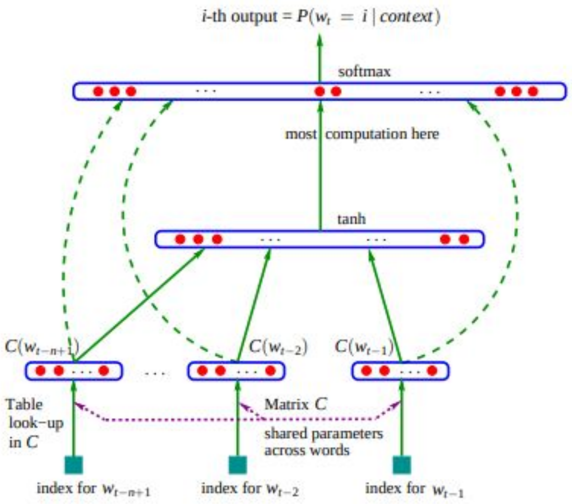
Neural Language Model

Neural Language Modeling:

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

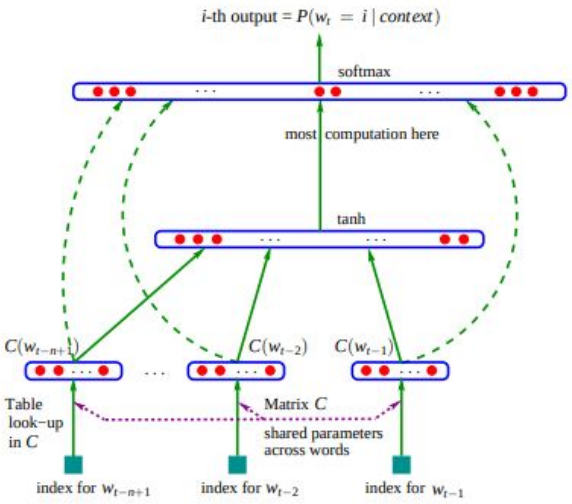


Neural Language Model

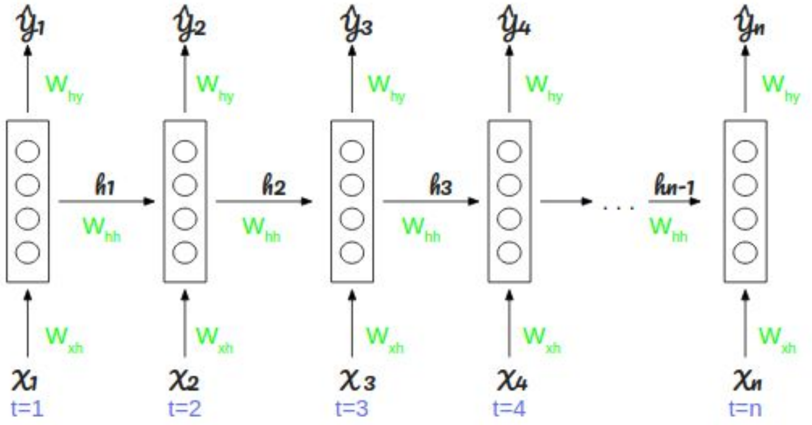


Feed Forward NN

Neural Language Model



Feed Forward NN



Recurrent NN

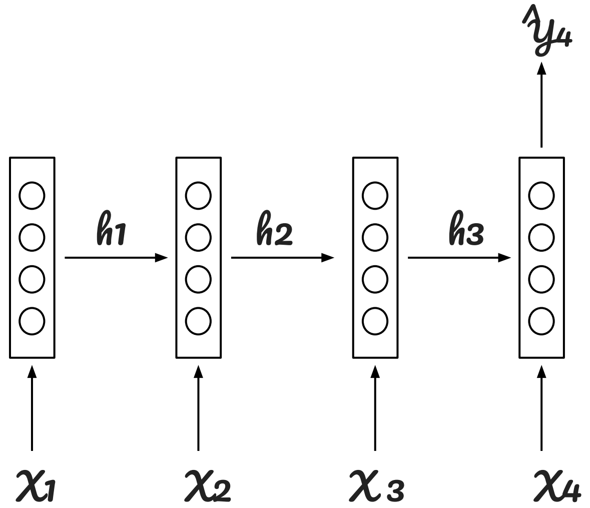
Analytics Vidhya
Learn everything about analytics

Neural Language Model

RNN for Language Modeling:

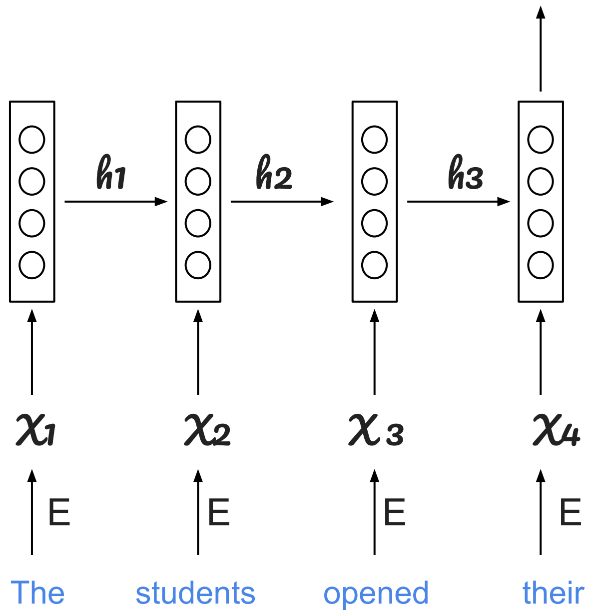
Neural Language Model

RNN for Language Modeling:



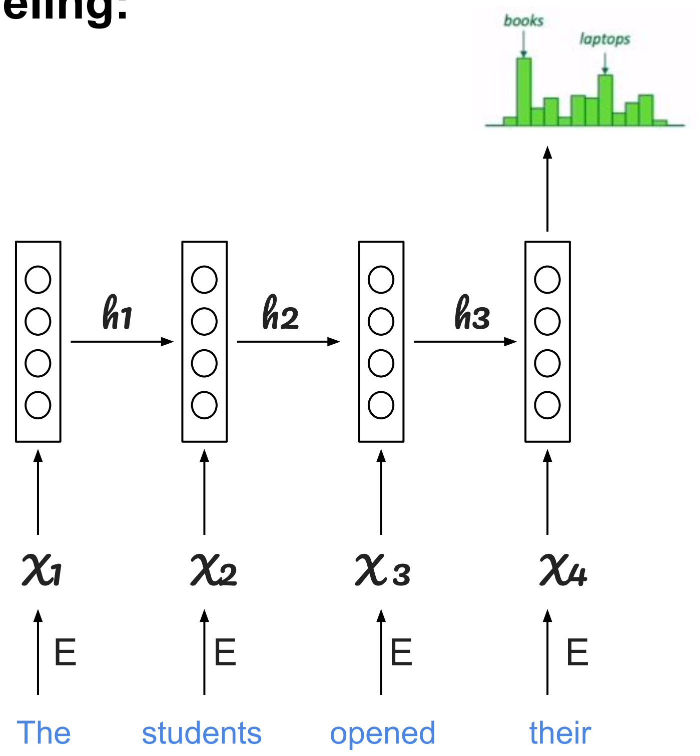
Neural Language Model

RNN for Language Modeling:



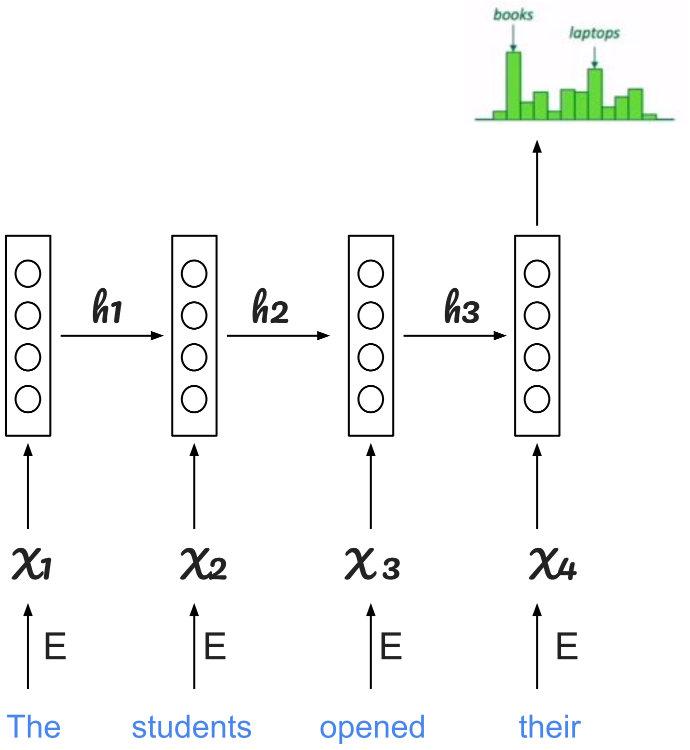
Neural Language Model

RNN for Language Modeling:



Neural Language Model

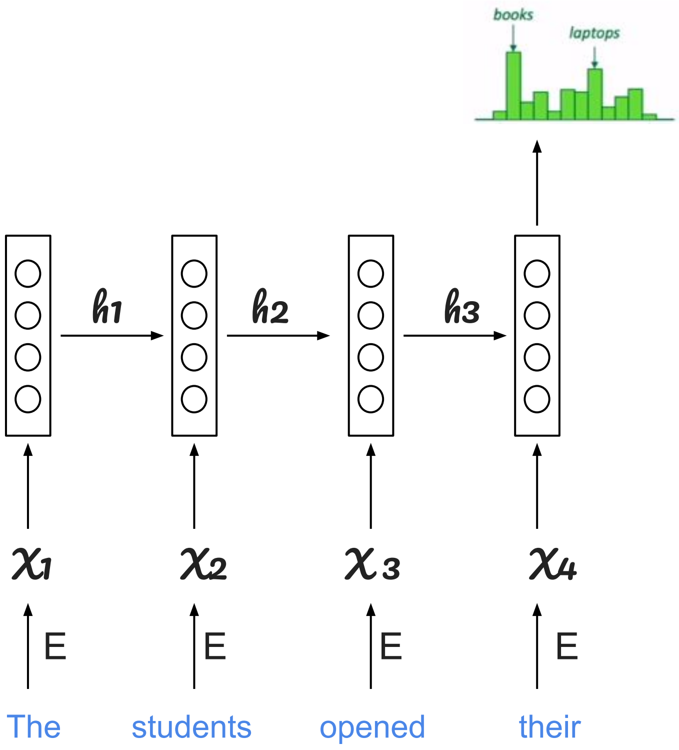
Advantages of RNN for Language Modeling:



Neural Language Model

Advantages of RNN for Language Modeling:

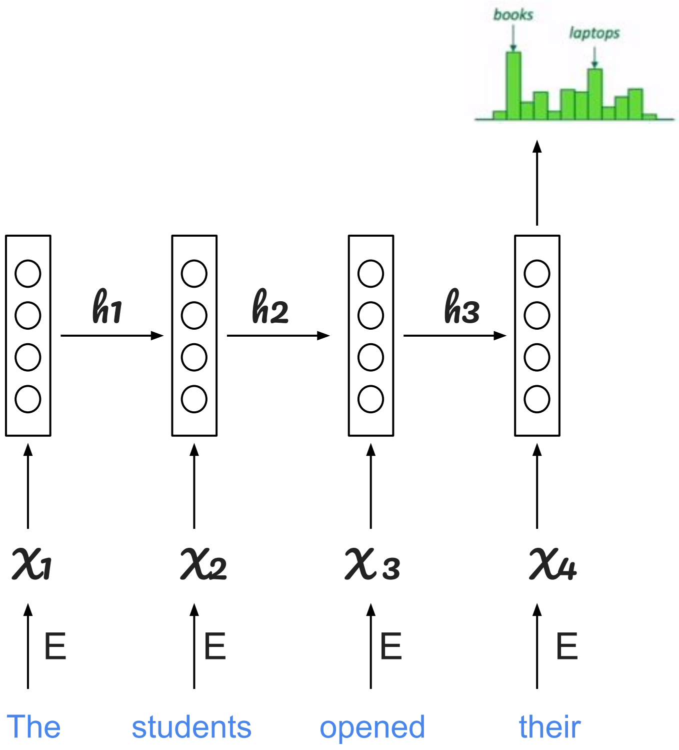
- Variable length input sequences



Neural Language Model

Advantages of RNN for Language Modeling:

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



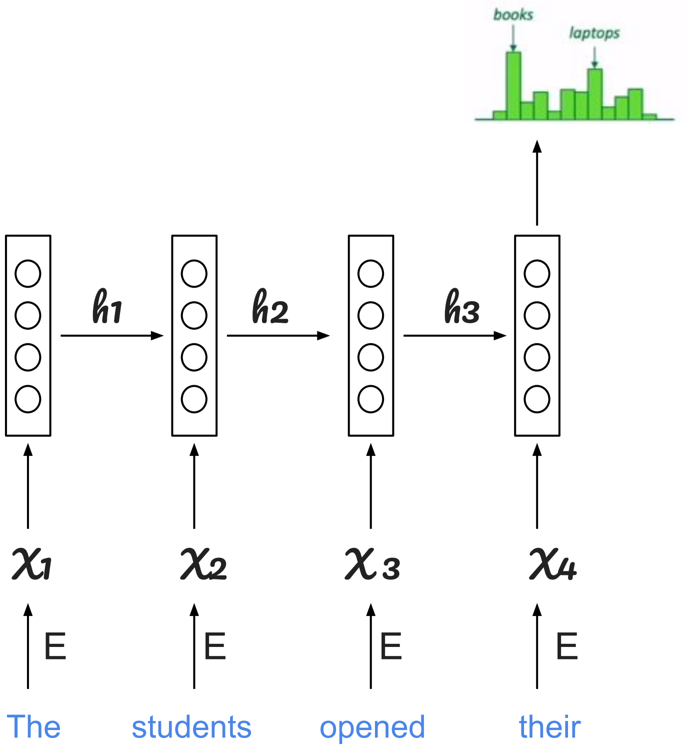
Neural Language Model

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>