Language Model

What is Language Model

A language model is a type of machine learning model trained to conduct a probability distribution over words.

For example,

I will go to _____

Which word is more appropriate here?

school.

market.

college.

These predictions can be done by a language model.

Types of language models

- Statistical language models
  - Statistical language models are a type of model that use statistical calculations in the data to make predictions about the likelihood of specific sequences of words.
    - Should It be, I will go to **school** or **market?**
  - A basic approach to building a probabilistic language model is to calculate **n-gram** probabilities.
- Neural language models
  - Neural language models use neural networks to predict the likelihood of a sequence of words.
  - These models are trained on a large corpus of text data and are capable of learning the underlying structure of the language.

N-gram model

- An n-gram model is a type of probabilistic language model used in natural language processing (NLP) to predict the next item in a sequence as a function of the preceding items.
- The "n" in n-gram represents the number of words in a given token sequence, and it can be adjusted based on the specific requirements of a task.

- So, if n = 1, then it will be unigram model (1-gram model)
- If n=2, then it will be bi-gram model (2-gram model)
- If n=3, then it will be a trigram-model (3-gram model)

And so on…

Recall some concepts

Conditional Probability:

$$P(B/A) = P(A \cap B)/P(A)$$

$$\text{Or, } P(A \cap B) = P(A) * P(B/A)$$

So, we can say,

$$P(A \cap B \cap C) = P(A) * P(B/A) * P(C/ A \cap B) \text{ ; Which is a chain rule}$$

We can write as below;

$$P(A \cap B \cap C \cap \ldots Z) = P(A) * P(B/A) * P(C/A \cap B) * P(D/A \cap B \cap C) \ldots\ldots$$

N-gram Model

Now, consider a sentence;

- I will go to school

P(**I** ∩ **will** ∩ **go** ∩ **to** ∩ **school**)

= P(**I**) * P(**will** / **I**) * P(**go** / **I** ∩ **will**) * P(**to** / **I** ∩ **will** ∩ **go**) * P(**school** / **I** ∩ **will** ∩ **go** ∩ **to**)

So, we can write it,

$P(w_1 \cap w_2 \cap w_3 \ldots w_{\square}) = \prod_i P(w_i / w_1 \cap w_2 \cap w_3 \ldots . w_{i-1})$

N-gram Model

So, we can write it,

$P(w_1 \cap w_2 \cap w_3 \ldots . w_n) = \prod_i P(w_i / w_1 \cap w_2 \cap w_3 \ldots . w_{i-1})$

For, Uni-gram: $P(w_1 \cap w_2 \cap w_3 \ldots . w_n) \approx \prod_i P(w_i)$

Bi-gram: $P(w_1 \cap w_2 \cap w_3 \ldots . w_n) \approx \prod_i P(w_i / w_{i-1})$

Tri-gram: $P(w_1 \cap w_2 \cap w_3 \ldots . w_n) \approx \prod_i P(w_i / w_{i-1} \cap w_{i-2})$

Four-gram: $P(w_1 \cap w_2 \cap w_3 \ldots . w_n) \approx \prod_i P(w_i / w_{i-1} \cap w_{i-2} \cap w_{i-3})$

N-gram Model

So, we can write it,

$P(w_1 \cap w_2 \cap w_3 \ldots . w_n) = \prod_i P(w_i / w_1 \cap w_2 \cap w_3 \ldots . w_{i-1})$

For, Uni-gram: $P(w_1 \cap w_2 \cap w_3 \ldots . w_n) \approx \prod_i P(w_i)$

Bi-gram: $P(w_1 \cap w_2 \cap w_3 \ldots . w_n) \approx \prod_i P(w_i / w_{i-1})$

Tri-gram: $P(w_1 \cap w_2 \cap w_3 \ldots . w_n) \approx \prod_i P(w_i / w_{i-1} \cap w_{i-2})$

Four-gram: $P(w_1 \cap w_2 \cap w_3 \ldots . w_n) \approx \prod_i P(w_i / w_{i-1} \cap w_{i-2} \cap w_{i-3})$

In practice, we usually use bi-gram, Tri-gram or Four-gram.

Bi-gram Model

The formula should be:

$$P(w_1 \cap w_2 \cap w_3 \ldots w_n) \approx \prod_i P(w_i / w_{i-1})$$

Here, $P(w_i / w_{i-1}) = P(w_i \cap w_{i-1}) / P(w_{i-1})$

Now, for a sentence;

I will go to ____ ; Next word might be **school** or **market**

$$P(\text{school/I will go to}) = \frac{N(\text{I will go to school})}{N(\text{I will go to})}$$

$$P(\text{market/I will go to}) = \frac{N(\text{I will go to market})}{N(\text{I will go to})}$$

Bi-gram Model

$$P(\text{school/I will go to}) = \frac{N(\text{I will go to school})}{N(\text{I will go to})}$$

N(I will go to school) = P(I / <S>) * P(will/ I) * P(go/ will) * P(to / go) * P(school / to)

N(I will go to) = P(I / <S>) * P(will/ I) * P(go/ will) * P(to / go)

So, P(school/I will go to) = P(school/ to)

Same, P(market/I will go to) = P(market/ to)

Tri-gram Model

The formula should be:

$$P(w_1 \cap w_2 \cap w_3 \ldots . w_{\square}) \approx \prod_i P(w_i / w_{i-1} \cap w_{i-2})$$

Here, $P(w_i / w_{i-1} \cap w_{i-2}) = P(w_i \cap w_{i-1} \cap w_{i-2}) / P(w_{i-1} \cap w_{i-2})$

Now, for the sentence;

I will go to ____ ; Next word might be **school** or **market**

$$P(\text{school/I will go to}) = \frac{N(\text{I will go to school})}{N(\text{I will go to})}$$

$$P(\text{market/I will go to}) = \frac{N(\text{I will go to market})}{N(\text{I will go to})}$$

Tri-gram Model

$$P(\text{school/I will go to}) = \frac{N(\text{I will go to school})}{N(\text{I will go to})}$$

N(I will go to school) = P(will/ <S> ∩ I) * P(go/ I ∩ will) * P(to / will ∩ go) * P(school / go ∩ to)

N(I will go to) = P(will/ <S> ∩ I) * P(go/ I ∩ will) * P(to / will ∩ go)

So, P(school/I will go to) = P(school/ go ∩ to)

Same, P(market/I will go to) = P(market/ go ∩ to)

Four-gram Model

$$P(\text{school/I will go to}) = \frac{N(\text{I will go to school})}{N(\text{I will go to})}$$

N(I will go to school) = P(go/ <S> ∩ I ∩ will) * P(to / I ∩ will ∩ go) * P(school / will ∩ go ∩ to)

N(I will go to) = P(go/ <S> ∩ I ∩ will) * P(to / I ∩ will ∩ go)

So, P(school/I will go to) = P(school/ will ∩ go ∩ to)

Same, P(market/I will go to) = P(market/ will ∩ go ∩ to)

N-gram Model

So, What would be the better way to predict the next word??

N-gram Model

So, What would be the better way to predict the next word??

If we consider all the previous words.

N-gram Model

So, What would be the better way to predict the next word??

If we consider all the previous words.

But there is a problem

Because the computation will be to complex

Exercise (Bi-gram)

Let us solve a small example to better understand the Bi-gram model.

Training corpus:

<S> I will go to the store.</S>

<S> I will go to the park.</S>

<S> She will go to school.</S>

<S> He will come here.</S>

<S> I will eat dinner.</S>

<S> I will play games.</S>

Test data:

<s> I will go to </s>

Next Word	Probability
P(the / to)	2/3
P(school / to)	1/3

Exercise (Bi-gram)

Let us solve a small example to better understand the Bi-gram model.

Training corpus:

<S> I will go to the store.</S>

<S> I will go to the park.</S>

<S> She will go to school.</S>

<S> He will come here.</S>

<S> I will eat dinner.</S>

<S> I will play games.</S>

Next Word	Probability
P(the / to)	2/3
P(school / to)	1/3

Test data:

<s> I will go to the </s> Here, the is more likeable.

Exercise (Tri-gram)

Let us solve a small example to better understand the Tri-gram model.

Training corpus:

<S>I will always choose the best option.</S>

<S>She will go to the market tomorrow.</S>

<S>He will make sure to call you.</S>

<S>I will go to the party tonight.</S>

<S>Will you go to the store?</S>

<S>I have to go to school tomorrow.</S>

<S>They will visit us next summer.</S>

<S>Tomorrow, I will go to the office early.</S>

<S>I need to go to the library this afternoon. </S>

Next Word	Probability
P(the / go to)	5/6
P(school / go to)	1/6

Test data: <s> I will go to </s>

Exercise (Tri-gram)

Let us solve a small example to better understand the Tri-gram model.

Training corpus:

<S>I will always choose the best option.</S>

<S>She will go to the market tomorrow.</S>

<S>He will make sure to call you.</S>

<S>I will go to the party tonight.</S>

<S>Will you go to the store?</S>

<S>I have to go to school tomorrow.</S>

<S>They will visit us next summer.</S>

<S>Tomorrow, I will go to the office early.</S>

<S>I need to go to the library this afternoon. </S>

Next Word	Probability
P(the / go to)	5/6
P(school / go to)	1/6

Test data: <s> I will go to the </s> Here, the is more likeable.

Exercise (Bi-gram)

Let check which sentence is more probable using Bi-gram model.

Training corpus:

<S> I will go to the store.</S>

<S> I will go to the park.</S>

<S> She will go to school.</S>

<S> He will come here.</S>

<S> I will eat dinner.</S>

<S> I will play games.</S>

Test data:

<S> I will play games </S>

<S> I will go to school </S>

Next Word	Probability
P(I / <S>)	4/6
P(will / I)	4/4
P(play / will)	1/6
P(games / play)	1/1

Exercise (Bi-gram)

Let check which sentence is more probable using Bi-gram model.

Training corpus:

<S> I will go to the store.</S>

<S> I will go to the park.</S>

<S> She will go to school.</S>

<S> He will come here.</S>

<S> I will eat dinner.</S>

<S> I will play games.</S>

Next Word	Probability
P(I / <S>)	4/6
P(will / I)	4/4
P(go / will)	3/6
P(to / go)	3/4
P(school / to)	1/3

Test data:

<s> I will play games </s> = (4/6) * (4/4) * (1/6) * (1/1) = 1/9

<S> I will go to school </S>

Exercise (Bi-gram)

Let check which sentence is more probable using Bi-gram model.

Training corpus:

<S> I will go to the store.</S>

<S> I will go to the park.</S>

<S> She will go to school.</S>

<S> He will come here.</S>

<S> I will eat dinner.</S>

<S> I will play games.</S>

Next Word	Probability
P(I / <S>)	4/6
P(will / I)	4/4
P(go / will)	3/6
P(to / go)	3/4
P(school / to)	1/3

Test data:

<s> I will play games </s> = (4/6) * (4/4) * (1/6) * (1/1) = 1/9

<S> I will go to school </S> = (4/6) * (4/4) * (3/6) * (3/4) * (1/3) = 1/12

Exercise (Bi-gram)

Let check which sentence is more probable using Bi-gram model.

Training corpus:

<S> I will go to the store.</S>

<S> I will go to the park.</S>

<S> She will go to school.</S>

<S> He will come here.</S>

<S> I will eat dinner.</S>

<S> I will play games.</S>

Next Word	Probability
P(I / <S>)	4/6
P(will / I)	4/4
P(go / will)	3/6
P(to / go)	3/4
P(school / to)	1/3

Test data:

<s> I will play games </s> = (4/6) * (4/4) * (1/6) * (1/1) = 1/9 [This sentence is more likely]

<S> I will go to school </S> = (4/6) * (4/4) * (3/6) * (3/4) * (1/3) = 1/12