Probabilistic Language Model

Probabilistic Language Model

- Task of assigning probability to a sentence or a phrase

  $P(S) = P(w_1, w_2, w_3, \ldots, w_n)$

- It can also be used to compute the probability of upcoming words

  $P(w_5 \mid w_1, w_2, w_3, w_4)$

Probabilistic Language Model

$$P(A|B) = P(A,B) / P(B)$$

Probabilistic Language Model

$$P(A|B) = P(A,B) / P(B)$$

$$P(A,B) = P(A|B) * P(B)$$

Probabilistic Language Model

$P(A|B) = P(A,B) / P(B)$

$P(A,B) = P(A|B) * P(B)$

$P(X_1, X_2, X_3, \ldots X_n) =$

Probabilistic Language Model

$$P(A|B) = P(A,B) / P(B)$$

$$P(A,B) = P(A|B) * P(B)$$

$$P(X_1,X_2,X_3,\ldots X_n) = P(X_1)$$

Probabilistic Language Model

$$P(A|B) = P(A,B) / P(B)$$

$$P(A,B) = P(A|B) * P(B)$$

$$P(X_1,X_2,X_3,\ldots X_n) = P(X_1) * P(X_2|X_1)$$

Probabilistic Language Model

$$P(A|B) = P(A,B) / P(B)$$

$$P(A,B) = P(A|B) * P(B)$$

$$P(X_1,X_2,X_3,\ldots X_n) = P(X_1) * P(X_2|X_1) * P(X_3|X_1,X_2)$$

Probabilistic Language Model

$$P(A|B) = P(A,B) / P(B)$$

$$P(A,B) = P(A|B) * P(B)$$

$$P(X_1,X_2,X_3,\ldots X_n) = P(X_1) * P(X_2|X_1) * P(X_3|X_1,X_2) * \ldots * P(X_n|X_1,X_2,\ldots,X_{n-1})$$

Probabilistic Language Model

can you come here

Probabilistic Language Model

can you come here

P(can,you,come,here) =

Probabilistic Language Model

can you come here

P(can,you,come,here) = P(can)

Probabilistic Language Model

can you come here

P(can,you,come,here) = P(can) * P(you|can)

Probabilistic Language Model

can you come here

P(can,you,come,here) = P(can) * P(you|can) * P(come|can,you)

Probabilistic Language Model

can you come here

P(can,you,come,here) = P(can) * P(you|can) * P(come|can,you) * P(here|can,you,come)

Probabilistic Language Model

can you come here

P(can,you,come,here) = P(can) * P(you|can) * P(come|can,you) * P(here|can,you,come)

P(here|can,you,come) = ?

Probabilistic Language Model

can you come here

P(can,you,come,here) = P(can) * P(you|can) * P(come|can,you) * P(here|can,you,come)

P(here|can,you,come) = Count of ("can you come here")

Probabilistic Language Model

can you come here

P(can,you,come,here) = P(can) * P(you|can) * P(come|can,you) * P(here|can,you,come)

P(here|can,you,come) = Count of (“can you come here”) / Count of (“can you come”)

Probabilistic Language Model

- What if there is no sentence “can you come here”

P(can,you,come,here) = P(can) * P(you|can) * P(come|can,you) * P(here|can,you,come)

P(here|can,you,come) = Count of (“can you come here”) / Count of (“can you come”)

Probabilistic Language Model

- What if there is no sentence “can you come here”
  - Count of (“can you come here) = 0

P(can,you,come,here) = P(can) * P(you|can) * P(come|can,you) * P(here|can,you,come)

P(here|can,you,come) = Count of (“can you come here”) / Count of (“can you come”)

Probabilistic Language Model

- What if there is no sentence “can you come here”
  - Count of (“can you come here) = 0
  - then P(here | can, you, come) = 0

P(can,you,come,here) = P(can) * P(you|can) * P(come|can,you) * P(here|can,you,come)

P(here|can,you,come) = Count of (“can you come here”) / Count of (“can you come”)

Probabilistic Language Model

$$P(w5 \mid w1, w2, w3, w4) = \frac{\text{Count of } (w1, w2, w3, w4, w5)}{\text{Count of } (w1, w2, w3, w4)}$$

Markov Assumption

$$P(w5 \mid w1, w2, w3, w4) = \frac{\text{Count of } (w1, w2, w3, w4, w5)}{\text{Count of } (w1, w2, w3, w4)}$$

- Markov Assumption: $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w3, w4)$

Markov Assumption

$$P(w5 \mid w1, w2, w3, w4) = \frac{\text{Count of } (w1, w2, w3, w4, w5)}{\text{Count of } (w1, w2, w3, w4)}$$

- Markov Assumption: $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w3, w4)$

- $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w4)$

Markov Assumption

$$P(w5 \mid w1, w2, w3, w4) = \frac{\text{Count of } (w1, w2, w3, w4, w5)}{\text{Count of } (w1, w2, w3, w4)}$$

- Markov Assumption: $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w3, w4)$
- $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w4)$
- $P(\text{here} \mid \text{can, you, come}) \approx P(\text{here} \mid \text{come})$

Markov Assumption

$$P(w5 \mid w1, w2, w3, w4) = \frac{\text{Count of } (w1, w2, w3, w4, w5)}{\text{Count of } (w1, w2, w3, w4)}$$

- Markov Assumption: $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w3, w4)$
- $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w4)$
- $P(\text{here} \mid \text{can, you, come}) \approx P(\text{here} \mid \text{come}) = \frac{\text{Count of (come here)}}{\text{Count of (come)}}$

Markov Assumption

$$P(w5 \mid w1, w2, w3, w4) = \frac{\text{Count of } (w1, w2, w3, w4, w5)}{\text{Count of } (w1, w2, w3, w4)}$$

- Markov Assumption: $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w3, w4)$
- $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w4)$
- $P(\text{here} \mid \text{can, you, come}) \approx P(\text{here} \mid \text{come}) = \frac{\text{Count of (come here)}}{\text{Count of (come)}}$
- $P(\text{here} \mid \text{can, you, come}) \approx P(\text{here} \mid \text{you, come})$

Markov Assumption

$$P(w5 \mid w1, w2, w3, w4) = \frac{\text{Count of } (w1, w2, w3, w4, w5)}{\text{Count of } (w1, w2, w3, w4)}$$

- Markov Assumption: $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w3, w4)$
- $P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w4)$
- $P(\text{here} \mid \text{can, you, come}) \approx P(\text{here} \mid \text{come}) = \frac{\text{Count of (come here)}}{\text{Count of (come)}}$
- $P(\text{here} \mid \text{can, you, come}) \approx P(\text{here} \mid \text{you, come}) = \frac{\text{Count of (you come here)}}{\text{Count of (you come)}}$

N-Gram Language Model

N-Gram Language Model

- An N-gram is a sequence of N tokens.

N-Gram Language Model

can you come here

- An N-gram is a sequence of N tokens.

N-Gram Language Model

can you come here

- An N-gram is a sequence of N tokens.

Unigrams:
N = 1, ["can", "you", "come", "here"]

N-Gram Language Model

can you come here

- An N-gram is a sequence of N tokens.

Unigrams:
N = 1, ["can", "you", "come", "here"]

$P(w5 \mid w1, w2, w3, w4) \approx P(w5)$

N-Gram Language Model

can you come here

- An N-gram is a sequence of N tokens.

Unigrams:
N = 1, ["can", "you", "come", "here"]

Bigrams:
N = 2, ["can you", "you come", "come here"]

N-Gram Language Model

can you come here

- An N-gram is a sequence of N tokens.

Unigrams:
N = 1, ["can", "you", "come", "here"]

Bigrams:
N = 2, ["can you", "you come", "come here"]

$$P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w4)$$

N-Gram Language Model

can you come here

- An N-gram is a sequence of N tokens.

Unigrams:
N = 1, [“can”, “you”, “come”, “here”]

Bigrams:
N = 2, [“can you”, “you come”, “come here”]

Trigrams:
N = 3, [“can you come”, “you come here”]

N-Gram Language Model

can you come here

- An N-gram is a sequence of N tokens.

Unigrams:
N = 1, ["can", "you", "come", "here"]

Bigrams:
N = 2, ["can you", "you come", "come here"]

Trigrams:
N = 3, ["can you come", "you come here"]

$$P(w5 \mid w1, w2, w3, w4) \approx P(w5 \mid w3, w4)$$

Thank You