

N gram model

① Language model

Sentence is assigned a probability.

~~Anurag ate~~

"Anurag killed a fish and ate it."

This makes it as a meaningful sentence,
Since words are arranged in a particular order, that makes sense.
If it would have been

"Killed fish Anurag and ate"

It does not make it a meaningful sentence.

→ This work is done by Language model
2 types.

① Joint probability.

② Conditional probability *** (for ^{imp and} n-grams)

If there are 2 words, we use this formula, and it determines, that after which word what word will come so that it makes sense

Teacher's Signature _____

$$P(x, y) \Rightarrow \frac{P(x|y) P(y)}{P(x)} \quad \text{or} \quad \frac{P(y|x) P(x)}{P(y)}$$

$$P(\text{google}, \text{it})$$

$$x = \text{google}$$

$$y = \text{it}$$

After google, there is it makes sense.

If More than 2 words,

$$P(x, y, z) = P(x) P(y|x) P(z|x, y)$$

$P(x) \Rightarrow$ probability of x

$P(y|x) \Rightarrow$ probability of y , when x is given.

$P(z|x, y) \Rightarrow$ probability of z when x, y are given.

$$P(\text{Hello, Miss Natasha}) = P(\text{Hello}) P(\text{Miss} | \text{Hello}) P(\text{Natasha} | \text{Hello Miss})$$

Markov Assumption.

I wish I was a unicorn.

$$P(\text{unicorn} \mid \text{I wish I was a})$$

$$\approx P(\text{unicorn} \mid a)$$

↑
one word before
unicorn

or else one / two words allowed
here.

(unigram, bi, tri, four, five, ... Ngram)

Bi | Trigram

bigram → checks the probability of
a word based on previous
word.

↖
Google it

Trigram → based on previous 2 words.

↖
Natasha Hello Mrs.

Teacher's Signature

unigram. (most easy)

- This happens contextually too.
- Ngram also detects the context of word.

For eg: I cried because she is dead.
↳ velt

The company name is cy.
eg. The cy foundation is a childcare ngo ← 1

my everyday because she was dead

N gram is also applied to the following:-
In Spelling correction :-

① Substitution \Rightarrow

language \Rightarrow language (considering context)

② Deletion

$P_{\text{neu}} \Rightarrow P_{\text{per}}$

③ Addition

Roast \Rightarrow Roasted.

④ Transposition

hangover \Rightarrow hangover
2 letters / 1 letter \rightarrow Arranged properly

Detects Saleam

I am honest - not

Bigram []

↑ if not detected

Trigram []

Not honest.

Ngram → Removes the contextual meaning in a broader spectrum of text.