Xavor AI Bootcamp

Week 1 lec 2

# LLM

Large Language Model

* **Generate sentence** from Set of Words
* **Predict next** word from Sequence of Words

Join Probability (Multiple properties, area, air speed, season)

# Probabilistic Language Models

Assign a probability to a sentence

E.g:

* Machine Translation
* Spell correction
* Speech Recognition

Capital W-set Small w-item of set

**Corpus-**

**P (B / A) = I am , I will , I had , I was , I should , I must**

**B=**  am, will, had, was , should, must , **A=** I

* *P(b/a)= p(a)p(b/a)*
* *P(a,b,c)= p(a)p(b/a)p(c/a,b)*

Multiply probability goes towards vanishing – chota hota jay ga (towards 0)

**P(b/a) = count (sentence with b) / count (sentence without a)**

**Markov Assumption**

*Current word Not dependent on all but dependent only on its previous word.*

# N-grams

Uni-gram , Bi -gram, tri-gram …. N-gram

Uni-gram : Independent

Bi-gram: Dependent on previous

Tri\_gram: Dependent on last 2 words

* Human- N-gram
* Computer- Uni-gram

# Estimating Bi-gram Probabilities

i= current word

i-1= previous words

P(wi|wi-1)= Count(wi-1, wi)/ Count(wi-1)

* *P(I|<s>)= I kitny dfa sentence Ky start ma ay ga if total sentences are 3 and I appear 2 times in start then 2/3=0.67,*
* *P(wi|wi-1)= Count(wi-1, wi)/ Count(wi-1)*

# Maximum Likelihood Classification

## Nave Bayes – 2 parts

1. Nave Bayes Classification

**Bag of words**

* Term (Word) Frequency
* No sequence Important

P(c/d)= [p(d/c)p(c)] / p(d)

P(c)= prior probability

P(d/c)= maximum likelihood

P(c/d)= posterial probability

1. ^ predicted and Calculated Value

C j= prior

Probability of cj (prior) = Count of Document having cj class / Total Documents

* Docunt(C=cj)/ N doc

Maximum Likley Hood

=count(wi, cj)/ summation count(w,cj)

= word in class/ word in dictionary

**Parameter Estimation**

Laplace (add+1)

=count(wi, cj)+1 / count((w,cj)+1)