

Q1) What is Language Modeling? Explain any one Language Modeling.

ANS. A Language Model (LM) is a system that assigns probability to a sequence of words.

It predicts:

- The next word in a sentence.
- Or the probability of a sentence being correct.

Example:

" I am going to the ____ "

→ Model predicts : market, school, park.

Applications:-

- Text Generation
- Machine Translation
- Speech Recognition
- Chatbots.

→ N-Gram Language Model:-

- An N-gram model predicts a word based on the previous (N-1) words.

• Formula:

$$P(w | w_{n-1}, \dots, w_1)$$

For bigram :

$$P(w | w_{n-1})$$

• Example (Bigram):

Sentence :

" I Love NLP "

Probability:

$$P(I) \times P(\text{love} | I) \times P(\text{NLP} | \text{Love})$$

Advantages :-

- Simple and Easy to implement.

Dis-Advantages :-

- Data Sparsity

- Cannot capture Long context.

→ Language Models help Machines Understands and generate Human Language.

Q2.) What is Transformer Model in NLP and How it Works?

ANS. The Transformer is a deep Learning Model used in NLP that relies on attention mechanism instead of recurrence. (RNN)

Key Idea :

Focus on important words in a sentence using attention.

Components :-

1. Input Embedding:

- Converts word into vectors.

2. Positional Encoding

- Adds information about word Order.

3. Self Attention Mechanism

- Each word attends to other words.

4. Multi-Head Attention

- Multiple Attention Layers learn different relationship.

5. Feed Forward Network

- Fully connected Layers.

6. Encoder - Decoder Structure

- Encoder - understands input

- Decoder - generates output.

• How it Works:-

1. Input sentence - embeddings.

2. Apply self Attention.

3. Capture relationship between words.

4. Generate Output.

- Advantages:

- ① Handles Long range dependencies.

- ② Parallel Processing. (faster than RNN)

- ③ High Accuracy.

- Used in:-

- ChatGPT

- BERT

- GPT Models.

Q3) What is Topic Modelling?

ANS.

Topic Modeling is an unsupervised learning technique used to discover hidden topics in a collection of documents.

Example:-

Document about:

- Sports

- Politics

- Technology.

Model Automatically groups them into topics.

Popular Methods:-

1. Latent Dirichlet Allocation (LDA)

- Most commonly used.

- Represents documents as mixture of topics.

2. Latent Semantic Analysis (LSA)

- Uses Matrix Decomposition.

⇒ How it works:-

1. Analyze word Frequency.

2. Identify patterns.

3. Group words into topics.

o Example:-

Topic: Sports

- Football, match, player.

Topic: Technology

- AI, Software, Data.

o Applications

1. Document Classification.

2. Recommendation systems.

3. Trend Analysis.

- Topic Modeling helps in Automatically discovering themes in text data.