MACHINE LEARNING PART 34

Natural Language Processing

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI) that focuses on the interaction between computers and human language.

It involves the development of algorithms and models to enable computers to understand, interpret, and generate human language.

Here are key concepts and techniques in NLP:

Tokenization

Definition:

- *Breaking text into smaller units, such as words or sentences (tokens).*

Importance:

- *Fundamental step for various NLP tasks.*

Part-of-Speech Tagging

Definition:

- *Assigning grammatical categories (e.g., noun, verb) to each word in a sentence.*

Importance:

- *Helps in understanding the syntactic structure of sentences.*

Named Entity Recognition

Definition:

- *Identifying and classifying entities (e.g., persons, organizations, locations) in text.*

Applications:

- *Information extraction, question answering.*

Word Embeddings

Definition:

- *Representing words as dense vectors in a continuous vector space.*

Techniques:

- *Word2Vec, GloVe, FastText.*

Applications:

- *Captures semantic relationships between words*

Sentiment Analysis

Definition:

- *Determining the sentiment expressed in a piece of text (positive, negative, neutral).*

Applications:

- *Social media monitoring, customer feedback analysis.*

Text Classification

Definition:

- *Assigning predefined categories or labels to text.*

Use Cases:

- *Spam detection, topic categorization.*

Language Modeling

Definition:

- *Predicting the probability of a sequence of words in a sentence.*

Applications:

- *Machine translation, speech recognition.*

Syntax and Parsing

Definition:

- *Analyzing the grammatical structure of sentences.*

Parsing Algorithms:

- *CKY, Earley, CYK.*

Applications:

- *Dependency parsing, syntactic analysis.*

Machine Translation

Definition:

- *Automatically translating text from one language to another.*

Applications:

- *Rule-based, statistical, neural machine translation.*

Question Answering

Definition:

- *Systems that answer questions posed in natural language.*

Techniques:

- *Information retrieval, machine comprehension.*

Speech Recognition

Definition:

- *Converting spoken language into written text.*

Techniques:

- *Hidden Markov Models, deep learning.*

Coreference Resolution

Definition:

- *Identifying when different expressions refer to the same entity.*

Applications:

- *Improves overall understanding of the text.*

Topic Modeling

Definition:

- *Identifying topics present in a collection of documents.*

Techniques:

- *Latent Dirichlet Allocation (LDA).*

Use Cases:

- *Document clustering, content recommendation.*

BERT

Definition:

- *Pre-trained transformer-based model for natural language understanding.*

Advantages:

- *Captures contextual information bidirectionally.*

Applications:

  - *Named entity recognition, sentiment analysis, question answering.*