Question Bank

Introduction to NLP

2-Marks Questions

- What is Natural Language Processing (NLP)?
- Mention two fields where NLP is used extensively.
- Define the role of NLP in customer service.
- What is tokenization in NLP?
- Define stemming in the context of text processing.
- What is lemmatization, and how does it differ from stemming?
- Why is it important to remove stopwords in text processing?
- Explain the purpose of case normalization in text preprocessing.
- How are regular expressions used in tokenization?

3-Marks Questions

- Discuss the application of NLP in the finance sector with examples.
- Describe the impact of NLP in healthcare with a specific use case.
- Explain the importance of NLP in modern-day applications.
- Compare and contrast stemming and lemmatization.
- Explain the process of tokenization at the word and sentence levels.
- Discuss the importance of handling punctuation during text preprocessing.
- Describe how removing stopwords can affect the results of text analysis.
- Explain the steps involved in basic text preprocessing in NLP.
- How can regular expressions be used to identify specific patterns in a text corpus?

4-Marks Questions

- Discuss the various applications of NLP across different fields such as healthcare, finance, and customer service.
- Provide a detailed explanation of the text preprocessing steps, including tokenization, stemming, lemmatization, stopword removal, and case normalization.
- Discuss the challenges and considerations involved in text preprocessing for NLP tasks.
- Compare the effectiveness of stemming and lemmatization with examples.
- Analyze the impact of proper text preprocessing on the performance of NLP models.
- Discuss the role of regular expressions in text preprocessing, including pattern matching and cleaning.

Explain how regular expressions can be used to automate complex text processing tasks with examples.

Linguistic Fundamentals for NLP

2-Marks Questions

- Define syntax in the context of NLP.
- What is a syntax tree? Provide a brief explanation.
- What are the primary goals of Part-of-Speech (POS) tagging?
- List two common POS tagging methods.
- What is Named Entity Recognition (NER) in NLP?
- Mention two applications of NER.
- Define dependency parsing in NLP.
- What are the common challenges faced in NER?

3-Marks Questions

- Explain the significance of parsing techniques in understanding sentence structure in NLP.
- Differentiate between syntax-based and dependency-based parsing.
- Discuss the use cases of POS tagging in real-world applications.
- Describe how POS tagging can be evaluated.
- Explain the role of NER in information extraction tasks.
- Discuss the challenges involved in accurately tagging parts of speech in a given text.
- Describe the process of building a syntax tree for a simple sentence.
- Discuss the common evaluation metrics used for NER systems.

4-Marks Questions

- Illustrate how POS tagging contributes to other NLP tasks, providing examples.
- Explain in detail the techniques used in NER and their respective challenges.
- Discuss how syntax and semantics work together in NLP to interpret the meaning of sentences.
- Describe the evaluation process for an NER system, including the metrics used and their significance.
- Explain how parsing techniques are used in machine translation and the challenges involved.

Machine Learning

2-Marks Questions

- What is supervised learning?
- Differentiate between classification and regression in supervised learning.
- Define precision in the context of model evaluation.

- What is recall, and how is it calculated?
- Briefly explain the F1-score and its importance in model evaluation.

- What is unsupervised learning?
- Define clustering in the context of unsupervised learning.
- Briefly describe the k-means clustering algorithm.
- What is hierarchical clustering?
- Define dimensionality reduction in machine learning.

- What is feature extraction in NLP?
- Define the Bag of Words model.
- What is TF-IDF, and how is it used in text analysis?
- Briefly explain the concept of word embeddings.
- Mention one difference between Word2Vec and GloVe.

- What is the Naive Bayes algorithm?
- Define Support Vector Machine (SVM) in the context of supervised learning.
- Briefly explain the k-Nearest Neighbors (k-NN) algorithm.
- What is a decision tree in machine learning?
- Mention one key advantage of using the Naive Bayes classifier.

3-Marks Questions

- Explain the key differences between classification and regression tasks in supervised learning.
- Describe a scenario where classification would be preferred over regression.

- Explain the key differences between supervised and unsupervised learning.
- Describe the process of the k-means clustering algorithm.
- Discuss the use of Principal Component Analysis (PCA) in dimensionality reduction.
- Explain how t-SNE is used for dimensionality reduction in high-dimensional data.

- Explain how the Bag of Words model works and its limitations.
- Discuss the importance of TF-IDF in feature extraction and its advantages over Bag of Words.
- Describe how Word2Vec generates word embeddings and their applications.
- Explain the differences between Word2Vec and GloVe in generating word embeddings.

- Describe how the Naive Bayes algorithm works with an example.
- Explain the working of the SVM algorithm and its application in classification tasks.
- Describe the structure of a decision tree and how it is used for making predictions.

4-Marks Questions

- Describe the process of supervised learning, including data preparation, model training, and evaluation.

- Explain how precision, recall, and F1-score are used together to evaluate the performance of a classification model.
- Compare and contrast the use of classification and regression in different real-world applications.
- Describe the working of the k-means clustering algorithm, including the steps involved in forming clusters.
- Compare k-means and hierarchical clustering, highlighting their strengths and weaknesses.
- Explain the importance of dimensionality reduction and PCA techniques used for this purpose.
- Discuss the applications of unsupervised learning in real-world scenarios.

- Compare and contrast Bag of Words and TF-IDF models for text feature extraction.
- Discuss the advantages of using word embeddings over traditional feature extraction methods like Bag of Words and TF-IDF.
- Explain in detail the process of generating word embeddings using Word2Vec.
- Discuss the applications of feature extraction techniques in NLP tasks like text classification and sentiment analysis.

- Explain in detail the steps involved in constructing a decision tree and how it is used for classification and regression tasks.
- Discuss the advantages and disadvantages of using SVMs compared to other classification algorithms.
- Describe the limitations of the k-NN algorithm and discuss strategies to overcome them.

Deep Learning for NLP

2-Marks Questions

- What is a neural network in the context of deep learning?
- Define an activation function and mention two common activation functions used in neural networks.
- What is the primary purpose of training a neural network?
- Briefly explain what Recurrent Neural Networks (RNNs) are.
- What problem does Long Short-Term Memory (LSTM) address in RNNs?
- What is a Transformer in the context of NLP?
- Define self-attention in the context of Transformers.
- Mention one application of BERT in NLP.

3-Marks Questions

- Explain the role of activation functions in neural networks.
- Describe the optimization process in training neural networks.
- Discuss the structure and working principle of Recurrent Neural Networks (RNNs).

- Explain how LSTM networks differ from standard RNNs and their advantage in sequence modeling.
- Describe the significance of self-attention in Transformers.
- Discuss how multi-head attention works in Transformer models.
- Explain the role of BERT in NLP tasks with an example.
- Discuss the main differences between RNNs and Transformers in processing sequential data.

4-Marks Questions

- Describe the basic structure of a neural network, including its components and how it functions.
- Explain in detail the training process of a neural network, including the role of backpropagation and gradient descent.
- Discuss the challenges faced by RNNs in handling long sequences and how LSTM networks address these issues.
- Illustrate the architecture of an LSTM cell and explain how it helps in retaining information over long sequences.
- Explain the architecture of a Transformer model, detailing the role of each component.
- Discuss the working of self-attention and multi-head attention in Transformers, providing examples of their applications in models like GPT.
- Describe the architecture and working principle of BERT, explaining how it differs from traditional NLP models.
- Compare and contrast the effectiveness of RNNs, LSTMs, and Transformers in NLP tasks, providing examples of their applications.

Advanced NLP Techniques

2-Marks Questions

- What is a language model in NLP?
- Mention two key features of the BERT architecture.
- Define text classification in the context of NLP.
- What is sentiment analysis, and why is it important?
- Briefly explain the purpose of Sequence-to-Sequence (seq2seq) models.
- Name two applications of seq2seq models in NLP.
- What distinguishes GPT from traditional language models?
- List one use case of RoBERTa in real-world applications.

3-Marks Questions

- Explain the architectural differences between BERT and GPT.
- Discuss the role of language models like RoBERTa in improving NLP tasks.
- Describe the process of text classification with an example.
- How does sentiment analysis work, and what are its common applications?
- Explain the architecture of a Sequence-to-Sequence model.

- Discuss the use of seq2seq models in machine translation with an example.
- Compare the effectiveness of BERT and GPT in handling different NLP tasks.
- Describe the challenges faced in sentiment analysis, particularly in handling sarcasm or irony.

4-Marks Questions

- Describe in detail the architecture of BERT and how it processes text differently from traditional models.
- Discuss the process of fine-tuning language models like BERT and RoBERTa for specific NLP tasks.
- Explain how text classification techniques are applied in sentiment analysis, including the use of machine learning algorithms.
- Illustrate the seq2seq architecture and discuss its application in chatbot development, including the handling of context in conversations.
- Discuss the role of attention mechanisms in improving the performance of Sequence-to-Sequence models.
- Compare and contrast the architectures of BERT, GPT, and RoBERTa, discussing their strengths and weaknesses in various NLP tasks.
- Explain the challenges and advancements in applying seq2seq models for summarization tasks, providing examples of current solutions.