

ASSIGNMENT NLP

1. What you understand by Text Processing? Write a code to perform text processing
2. What you understand by NLP toolkit and spacy library? Write a code in which any one gets used.
3. Describe Neural Networks and Deep Learning in Depth
4. what you understand by Hyperparameter Tuning?
5. What you understand by Ensemble Learning?
6. What do you understand by Model Evaluation and Selection ?
7. What you understand by Feature Engineering and Feature selection? What is the difference between them?

1. Text processing involves manipulating and analyzing text data to extract useful information. Here's a simple code snippet in Python to perform basic text processing tasks like tokenization and stemming using the NLTK library:

CODE:

```
import nltk
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer

nltk.download('punkt')

# Sample text
text = "Text processing involves analyzing and manipulating text data."

# Tokenization
tokens = word_tokenize(text)
print("Tokens:", tokens)

# Stemming
stemmer = PorterStemmer()
stemmed_tokens = [stemmer.stem(token) for token in tokens]
print("Stemmed Tokens:", stemmed_tokens)
```

2.NLP toolkit like NLTK and spaCy provide tools and resources for natural language processing tasks such as tokenization, part-of-speech tagging, named entity recognition, and more. spaCy is a more modern and efficient library compared to NLTK. Here's a code example using spaCy:

```
import spacy

# Load the English language model
nlp = spacy.load('en_core_web_sm')

# Sample text
text = "Natural language processing (NLP) is a field of AI."

# Process the text
doc = nlp(text)

# Print token and POS tags
for token in doc:
    print(token.text, token.pos_)
```

3.Neural networks are computational models inspired by the human brain's structure and function, capable of learning complex patterns from data. Deep learning is a subfield of machine learning that utilizes neural networks with multiple hidden layers to automatically learn features from data. These networks have shown remarkable success in various tasks such as image recognition, natural language processing, and more.

4.Hyperparameter tuning involves optimizing the hyperparameters of a machine learning model to improve its performance. This process typically involves techniques like grid search, random search, or more advanced optimization algorithms to find the best combination of hyperparameters for a given model and dataset.

5.Ensemble learning is a machine learning technique where multiple models are combined to improve the overall performance. It involves training several base models and then combining their predictions to make a final prediction. Common ensemble methods include bagging (e.g., Random Forest), boosting (e.g., Gradient Boosting Machines), and stacking.

6.Model evaluation and selection refer to the process of assessing the performance of machine learning models and choosing the best-performing model for deployment. It involves techniques

such as cross-validation, evaluation metrics (e.g., accuracy, precision, recall, F1-score), and comparison of different models based on their performance on validation or test datasets.

7. Feature engineering involves creating new features from existing data to improve the performance of machine learning models, while feature selection involves choosing a subset of relevant features to improve model efficiency and interpretability. Feature engineering focuses on creating meaningful representations of data, while feature selection aims to reduce the dimensionality of the data while preserving important information.