Bytes of Intelligence: Set - 06

Department of AI Engineering

Semester Examination

Course Title: Deep Learning Techniques in Python for Vision and Language

Course Code: DL-101

Duration: 3 Hours

Total Marks: 100

Instructions:

- **Attempt all questions.**
- **Read each question carefully before answering.**
- **All code should be properly commented and indented.**
- **Use appropriate data structures and algorithms where necessary.**
- **Marks are indicated next to each question.**
- **Assume any necessary imports (e.g., `import numpy as np`, `import pandas as pd`).**
- **Write your code in Python programming language.**

Instructions:

- **Attempt all questions.**
- **Read each question carefully before answering.**
- **All code should be properly commented and indented.**
- **Use appropriate data structures and algorithms where necessary.**
- **Marks are indicated next to each question.**
- **Assume any necessary imports (e.g., `import numpy as np`, `import pandas as pd`).**
- **Write your code in Python programming language.**

Question 1: Python Fundamentals and Control Structures (15 Marks)

(a) Basic Data Types and Operations (5 Marks)

Write a Python program that:

- Prompts the user to input a string containing a mathematical expression (e.g., `"3 + 4 * 2"`).
- Evaluates the expression and prints the result.
- Handles any exceptions (e.g., `SyntaxError`, `ZeroDivisionError`) and prints an appropriate error message.

Example:

```
Enter a mathematical expression: 3 + 4 * 2
Result: 11
```

(b) Control Flow and Functions (10 Marks)

Define a function `fibonacci_sequence(n)` that generates a list containing the Fibonacci sequence up to the `n` th term.

- The function should return the list of Fibonacci numbers.
- Include error handling to ensure that `n` is a positive integer.
- Demonstrate the function with `n = 10`.

Question 2: Data Structures in Python (20 Marks)

(a) List and Dictionary Manipulations (10 Marks)

Given the following list of dictionaries representing students and their test scores:

```
students = [
    {"name": "Alice", "score": 88},
    {"name": "Bob", "score": 75},
    {"name": "Charlie", "score": 92},
    {"name": "Diana", "score": 85},
    {"name": "Evan", "score": 79}
]
```

Perform the following tasks:

- Sort the list of students by their scores in ascending order. *(5 Marks)*
- Create a dictionary where each key is a student's name and the value is their score. *(5 Marks)*

(b) Set Operations (10 Marks)

You have two sets representing programming languages known by two developers:

```
developer_1 = {"Python", "Java", "C++", "JavaScript"}
developer_2 = {"Ruby", "JavaScript", "Python", "Go"}
```

Perform the following tasks:

- Find the set of languages known by both developers. *(3 Marks)*
- Find the set of languages known by developer_1 but not by developer_2. *(3 Marks)*
- Create a set of all languages known by either developer. *(4 Marks)*

Question 3: NumPy Array Operations (15 Marks)

(a) Array Creation and Mathematical Operations (10 Marks)

- Create a NumPy array `A` of shape `(3, 3)` with values from 1 to 9. *(2 Marks)*
- Compute the inverse of matrix `A`. *(4 Marks)*
- Verify that the dot product of `A` and its inverse yields the identity matrix. *(4 Marks)*

Note: Handle the case where the matrix might be singular.

(b) Statistical Functions (5 Marks)

Given a NumPy array of exam scores:

```
scores = np.array([88, 75, 92, 85, 79, 95, 68, 74, 81, 90])
```

- Calculate the mean, median, and standard deviation of the scores. *(3 Marks)*
- Determine the percentage of students who scored above the mean. *(2 Marks)*

Question 4: Data Analysis with Pandas (25 Marks)

You are provided with a CSV file `transactions.csv` containing the following columns: `'TransactionID'`, `'CustomerID'`, `'Product'`, `'Category'`, `'Quantity'`, `'UnitPrice'`, `'TransactionDate'`.

(a) Data Loading and Cleaning (10 Marks)

- Load the dataset into a Pandas DataFrame. *(2 Marks)*

- Convert `'TransactionDate'` to datetime format and extract the month into a new column `'Month'`. *(3 Marks)*
- Check for missing values and handle them appropriately. Explain your method. *(5 Marks)*

(b) Data Aggregation and Grouping (10 Marks)

- Calculate the total revenue for each product (`Quantity` * `UnitPrice`). *(3 Marks)*
- Identify the top 5 products by total revenue. *(4 Marks)*
- Group the data by `'Category'` and calculate the total quantity sold for each category. *(3 Marks)*

(c) Data Visualization (5 Marks)

- Using Matplotlib or Seaborn, create a bar chart showing the total revenue for the top 5 products identified in part (b).
- Include appropriate labels, title, and color customization.

Question 5: Data Visualization with Matplotlib and Seaborn (15 Marks)

(a) Time Series Visualization (7 Marks)

Using the `transactions.csv` DataFrame:

- Calculate the monthly total revenue.
- Plot a line chart showing the monthly total revenue over time.
- Customize the plot with labels for the x-axis (months), y-axis (total revenue), a title, and gridlines.

(b) Advanced Visualization (8 Marks)

- Use Seaborn to create a box plot of the unit prices for each product category.
- Interpret any notable differences in unit price distributions across categories.

Question 6: Hands-On Project (15 Marks)

You are given a dataset `employee_attendance.csv` containing the following columns: `'EmployeeID'`, `'Date'`, `'Status'` (Present/Absent), `'Department'`.

Tasks:

- Load the dataset into a Pandas DataFrame and parse the `'Date'` column as datetime. *(3 Marks)*

- Calculate the attendance percentage for each employee. *(4 Marks)*
- Identify the department with the highest average attendance. *(4 Marks)*
- Using Matplotlib, create a bar chart showing the top 5 employees with the highest attendance percentages. Include labels and title. *(4 Marks)*

Question 7: Capstone Coding Challenge (Bonus Question - Optional) (10 Marks)

Data Processing and Sentiment Analysis

Write a Python script that:

- Reads a text file `'reviews.txt'` containing customer reviews, one review per line. *(1 Mark)*
- Cleans the text by removing punctuation, converting to lowercase, and tokenizing into words. *(2 Marks)*
- Counts the frequency of positive and negative words using the provided lists `positive_words` and `negative_words`. *(3 Marks)*

```
positive_words = ['good', 'great', 'excellent', 'amazing', 'fantastic', 'love']
negative_words = ['bad', 'poor', 'terrible', 'awful', 'hate', 'worst']
```

- Calculates the overall sentiment score (positive count minus negative count). *(2 Marks)*
- Plots a pie chart showing the proportion of positive and negative words in the reviews. *(2 Marks)*