

## Lab-1&2

### CSL2050 - Pattern Recognition and Machine Learning

#### **NOTE:**

1. This Programming Assignment is associated with your Lab-1 and Lab-2. Part of the problems you are recommended to complete in lab hours and remaining is take home.
2. Do this assignment in Google Colab with proper comments and notes. High-quality documentation is expected for full marks.
3. Submit single *PA1\_rollNum.ipynb* where rollNum is your roll number. Any violation in file naming may lead to 0 marks. For this assignment, there is no requirement for a separate report.
4. Please refer to the Academic Code of Honor for this course (ref: Lecture-1) before submission of this programming assignment. Additionally, you are not allowed to use LLM for this assignment.
5. **Deadline:** January 20, 2023, 10:30 PM.
6. **Late Submission:** Late submissions beyond the due date will incur a 10% penalty for each day. Plan the submission ahead and do not wait for last minute.

---

**Practice:** Please go through the following tutorial before attempting the problems.

1. Numpy Tutorial: <https://cs231n.github.io/python-numpy-tutorial/>
2. Pandas Tutorial <https://www.w3schools.com/python/pandas/default.asp>
3. HTML Parser: <https://docs.python.org/3/library/html.parser.html>

#### **Problems:**

1. **Lists and Loops:** Write a Python function that takes a list of integers as input and returns the sum of all even numbers in the list. **(2 points)**
2. **nearest neighbour:** Given a 2D-point and an array of 2D-points, find out three nearest neighbour. **(2 points)**
3. **List Comprehension:** Given a list of words, create a new list containing the length of each word using list comprehension. **(2 points)**
4. **NumPy Array Operations:** Given two NumPy arrays `arr1 = np.array([1, 2, 3])` and `arr2 = np.array([4, 5, 6])`, perform element-wise multiplication and store the result in a new array. **(2 points)**
5. **NumPy Array Slicing:** Given a NumPy array `data = np.arange(1, 21)`, use array slicing to extract elements from index 5 to index 15. **(2 points)**
6. **Data Visualization with Matplotlib:** Using Matplotlib, plot the following functions: (a)  $y = 5x + 4$  where  $x \in [-1, 1]$ , (b)  $y = \ln(x)$  where  $x > 10$  and  $x < 100$ , (c)  $y = x^2$  where  $x \in [-10, 10]$  (d)  $y = \frac{1}{1+e^x}$  where  $x \in [-1, 1]$ . **(5 points)**
7. **Data Cleaning with Pandas:** Given a Pandas DataFrame with a column containing missing values, write a Python function to replace those missing values with the mean of that column. **(5 points)** [Data: data/feature.csv]
8. **Pandas DataFrame Filtering:** Read the CSV file and using Pandas, filter a DataFrame to only include rows where the 'age' column is greater than 40 and the 'gender' column is 'female'. **(5 points)** [Data: data/demo.csv]
9. **Pandas Grouping:** Given a data frame with three columns, 'Country,' 'City', and 'Population,' use Pandas to group the data by 'Country' and calculate each country's average population for cities. **(5 points)** [Data: data/population.csv]
10. **Data Manipulation with Pandas:** Given a Pandas DataFrame containing columns 'Sales' and 'Expenses', create a new column 'Profit' that calculates the profit as 'Sales' minus 'Expenses'. You may take your own data and demonstrate its working. **(5 points)**
11. **Dictionary Manipulation:** Given a dictionary phonebook containing names as keys and phone numbers as values, write a function to find and return the name(s) of the person(s) with the maximum phone number(s). You may create phonebook of 10 of your friends. **(5 points)**

12. **Function** Write a Python function that takes a vector as input and computes ReLU activation on each element of that vector. Note:  $\text{ReLU}(x) = \max(0, x)$ . **(5 points)**
13. **File Handling:** Read a text file named “data.txt” that contains one number per line. Write a Python script to calculate the sum of all the numbers and print the result. **(5 points)** [Data: data/marks.csv]
14. **Object-Oriented Programming:** Create a class Rectangle with attributes width and height, and methods to calculate the area and perimeter. Test the class by creating objects and performing calculations. **(5 points)**
15. **Data Visualization with Matplotlib:** Using Matplotlib, create a bar chart to compare students’ average scores and variance in scores in five subjects. **(5 points)** [Data: data/marks.csv]
16. Data crawling is an important aspect of Machine Learning. You are given the following website: <https://openaccess.thecvf.com/ICCV2023?day=all>. Write a Python script that gets a CSV file containing author names, the title of the paper, and PDF as the columns. **(10 points)**

End of Paper