MDS5102 Python Programming

Assignment 2

Due Date: 28 November, 2021

Assignment Description:

This assignment will worth **10%** of the final grade.

You should write your code for each question in a .py file (please name it using the question name, e.g. q1.py). Please pack all your python code into a single .zip file, name it using your student ID (e.g. if your student ID is 123456789 and to submit Assignment 2, then the file should be named as 123456789_Assignment2.zip or 123456789_A2.zip), and then **submit the .zip** file via BlackBoard.

Please also write a report file, which provide the details of your codes. (Note that the report should be submitted as PDF). The report should also be included in the .zip file as well.
Please note that, the teaching assistant may ask you to explain the meaning of your program, to ensure that the codes are indeed written by yourself. Plagiarism will not be tolerated. We may check your code using Blackboard.

This assignment is due on **23:59PM, 28 Nov (Sunday)**. For each day of late submission, you will lose 10% of your mark in corresponding assignment. If you submit more than three days later than the deadline, you will receive zero in this assignment.

Question 1 (4% × 5 = 20% of this assignment)

Write a program that allow user to input a positive integer N,

(a) Create a N × N matrix that value ranging from 1 to N^2, positioning from upper left to lower right.
e.g. N=3, the matrix should be [[1, 2, 3], [4, 5, 6], [7, 8, 9]]

(b) Compute the determinant of matrix created in (a)

(c) The inverse of matrix is another matrix, which on multiplication with the given matrix gives the multiplicative identity. For a matrix A, its inverse is A^{-1}, and $A \cdot A^{-1} = I$. Find the inverse matrix of the matrix created in (a)

(d) Create a N × N identity matrix, validate whether the result in (c) satisfies definition of inverse matrix. Print True if so and False otherwise.

(e) Define a function whose input is a matrix and return normalized input. And use the defined function normalize the result in (a) and (c)

Question 2 (10% × 3 = 30% of this assignment)

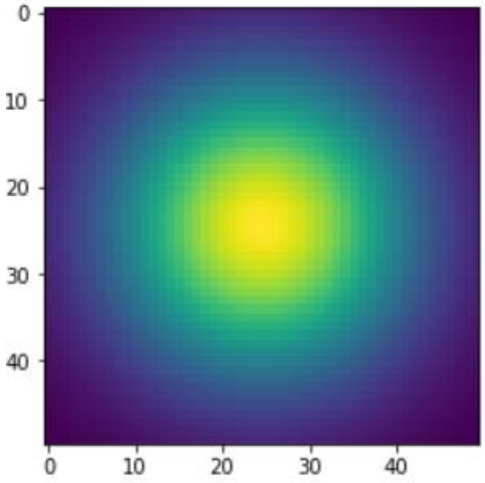
(a) **2D Gaussian Matrix** Write a program that allow user to input a number s and positive integer N, generate a generic 2D Gaussian-like array G.
Here we only consider the area whose horizontal and vertical coordinates are between -1 and 1 and end point included in cartesian coordinate system. ($x \in [-1, 1], y \in [-1, 1]$). The result array has resolution of N × N, and its mean is zero while standard deviation is s.
Mathematically, the 2D Gaussian kernel is computed as

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

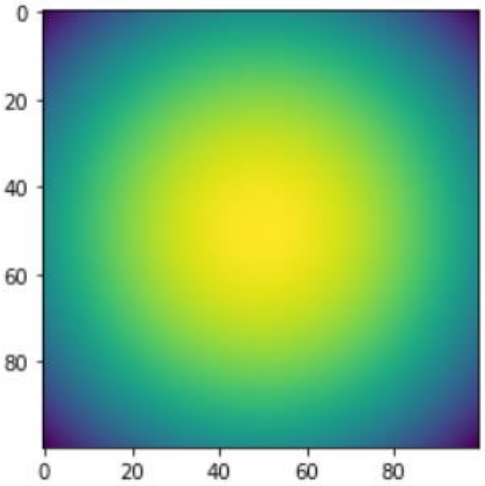
Plot the 2D array using the following code:

```
import matplotlib.pyplot as plt
plt.imshow(G)
```

Example:
s=1.0, N=50



s=50.0, N=100



(b) **Moving Average** Write a function that takes 2 input arguments, one is a 1-dimensional numpy array *arr* and the other is a positive integer *N*, output of this function is an array of moving averages of *arr*, using a sliding window with length of *N*. Suppose length of input array is n, then the length of result would be $n - N + 1$.

Example:
src = np.range(15)
moving_average(src, 3)

```
[ 1.  2.  3.  4.  5.  6.  7.  8.  9. 10. 11. 12. 13.]
```

moving_average(src, 6)

```
[ 2.5  3.5  4.5  5.5  6.5  7.5  8.5  9.5 10.5 11.5]
```

(c) **Matrix Creation** Write a program that requires user to input a positive odd integer N, create a N×N matrix satisfies that:

(1) Element at center always equals to N, and is surrounded by N-1.
(2) (N-1)s is further surrounded by (N-2)s, and (N-2)s are further surrounded by (N-3)s.
(3) Loops like this until the outmost of the N×N matrix.

Numpy package shall be used and you can assume that input is always a positive number. Output examples shown as following:

Input: 7

Output:

```
4 4 4 4 4 4 4
4 5 5 5 5 5 4
4 5 6 6 6 5 4
4 5 6 7 6 5 4
4 5 6 6 6 5 4
4 5 5 5 5 5 4
4 4 4 4 4 4 4
```

Input: 1

Output:

```
1
```

Question 3 (3%×10=30% of this assignment)

Given the csv file in attachment *q3.csv*, write code to complete following actions. Your code must be presented in submission.

(a) Read the csv file as a data frame.
(b) Present number of observations in the dataset. Present number of columns in the dataset. Print the name of all columns.
(c) Present the first 10 entries of the data. Then present last 15 entries.
(d) Present Education with least occurrence.
(e) Present the number of Year_Birth in the dataset.
(f) Present the most frequent Marital_Status.
(g) Present list all values of Kidhome of the most frequent Marital_Status.
(h) Present the min and max value of Income of different Education levels.
(i) Present mean values and standard deviation of MntWines (Mount of Wines) in different Marital_Status.
(j) Based on statistics in part (i), describe the relationship between MntWines and Marital_Status. You can also present more data to analyze.

Question 4 (4%×5=20% of this assignment)

Read data in *q4.txt*, which is a dataset of occupation. Complete following tasks.

(a) Read data as a data frame indexed by user_id.
(b) Present the mean and variance of age per occupation.
(c) Calculate the Female ratio per occupation and sort it from the most to the least.
(d) Calculate the mean age for each combination of occupation and gender.
(e) Calculate the percentage of Male and Female for each occupation.

End of Assignment