UIT2502---Data Analytics and Visualization Lab

Ex 1 a: Basic Numpy Exercise

Name: Vasundhara.B

Roll No: 3122 21 5002 119

1. Import numpy as np and see the version

CODE:

```
C: > Users > 3122215002119 > Documents > EX_1a.py
1   import numpy as np
2   print(np.__version__)
```

OUTPUT:

```
1.25.2
PS C:\Users\3122215002119\Documents>
```

2. How to create a 1D array?

CODE:

```
C: > Users > 3122215002119 > Documents > EX_1a.py > ...
1   import numpy as np
2   list = [1, 2, 3, 4, 5]
3   n = np.array(list)
4   print(n)
```

OUTPUT:

```
[1 2 3 4 5]
PS C:\Users\3122215002119\Documents>
```

3. How to create a boolean array?

CODE:

```
C: > Users > 3122215002119 > Documents > EX_1a.py > ...
1   import numpy as np
2
3   arr=[5, None, 1, 25, -10, 0, 'A']
4   bool_arr = np.array(arr, dtype='bool')
5   print(f'Boolean Array: {bool_arr}')
```

OUTPUT:

```
Boolean Array: [ True False  True  True  True False  True]
PS C:\Users\3122215002119\Documents>
```

4. How to extract items that satisfy a given condition from 1D array?

CODE:

```
C: > Users > 3122215002119 > Documents > EX_1a.py > ...
1   import numpy as np
2
3   arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
4   print(arr[arr%2 == 1])
```

OUTPUT:

```
[1 3 5 7 9]
PS C:\Users\3122215002119\Documents>
```

5. How to replace items that satisfy a condition with another value in numpy array?

CODE:

```
C: > Users > 3122215002119 > Documents > EX_1a.py > ...
1   import numpy as np
2
3   arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
4   arr[arr%2 == 1] = -1
5   print(arr)
```

OUTPUT:

```
[ 0 -1  2 -1  4 -1  6 -1  8 -1]
PS C:\Users\3122215002119\Documents>
```

6. How to replace items that satisfy a condition without affecting the original array?

CODE:

```
C: > Users > 3122215002119 > Documents > EX_1a.py > ...
1   import numpy as np
2
3   arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
4   out = arr.copy()
5   out[out%2 == 1] = -1
6
7   print('Modified Array')
8   print(out)
9
10  print('\nOriginal Array')
11  print(arr)
```

OUTPUT:

```
Modified Array
[ 0 -1  2 -1  4 -1  6 -1  8 -1]

Original Array
[0 1 2 3 4 5 6 7 8 9]
PS C:\Users\3122215002119\Documents>
```

7. How to reshape an array?

CODE:

```
C: > Users > 3122215002119 > Documents > EX_1a.py > ...
1   import numpy as np
2
3   arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
4   narr=arr.reshape(2, -1)
5   print(narr)
```

OUTPUT:

```
[[0 1 2 3 4]
 [5 6 7 8 9]]
PS C:\Users\3122215002119\Documents>
```

8. How to stack two arrays vertically?

CODE:

```
C: > Users > 3122215002119 > Documents > EX_1a.py > ...
1   import numpy as np
2
3   arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
4   a = arr.reshape(2,-1)
5   b = arr.reshape(2,-1)
6   sarr=np.vstack([a,b])
7   print(sarr)
```

OUTPUT:

```
[[0 1 2 3 4]
 [5 6 7 8 9]
 [0 1 2 3 4]
 [5 6 7 8 9]]
PS C:\Users\3122215002119\Documents>
```

9. How to stack two arrays horizontally?

CODE:

```
import numpy as np

arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
a = arr.reshape(2, -1)
b = arr.reshape(2, -1)
sarr = np.hstack([a,b])
print(sarr)
```

OUTPUT:

```
[[0 1 2 3 4 0 1 2 3 4]
 [5 6 7 8 9 5 6 7 8 9]]
```

10. How to generate custom sequences in numpy without hardcoding?

CODE:

```
import numpy as np

arr = np.array([1, 2, 3])
res1 = np.r_[np.repeat(arr, 3)]
print(res1)
res2 = np.r_[np.tile(arr, 3)]
print(res2)
```

OUTPUT:

```
[1 1 1 2 2 2 3 3 3]
[1 2 3 1 2 3 1 2 3]
```

11. How to get the common items between two python numpy arrays?

CODE:

```
import numpy as np

a = np.array([1,2,3,2,3,4,3,4,5,6])
b = np.array([7,2,10,2,7,4,9,4,9,8])
c = np.intersect1d(a,b)
print(c)
```

OUTPUT:

```
[2 4]
```

12. How to remove from one array those items that exist in another?

CODE:

```
import numpy as np

a = np.array([1,2,3,4,5])
b = np.array([5,6,7,8,9])
c = np.setdiff1d(a,b)
print(c)
```

OUTPUT:

```
[1 2 3 4]
```

13. How to get the positions where elements of two arrays match?

CODE:

```
import numpy as np

a = np.array([1,2,3,2,3,4,3,4,5,6])
b = np.array([7,2,10,2,7,4,9,4,9,8])
c = np.where(a == b)
print(c[0])
```

OUTPUT:

```
[1 3 5 7]
```

14. How to extract all numbers between a given range from a numpy array?

CODE:

```
import numpy as np

a = np.array([2, 6, 1, 9, 10, 3, 27])
print(a[(a >= 5) & (a <= 10)])
```

OUTPUT:

```
[ 6  9 10]
```

15. How to make a python function that handles scalars to work on numpy arrays?

CODE:

```python
import numpy as np

def maxx(x, y):
    if x >= y:
        return x
    else:
        return y

def pair_max(x, y):
    maximum = [maxx(a,b) for a,b in map(lambda a,b:(a,b),x,y)]
    return np.array(maximum)

a = np.array([5, 7, 9, 8, 6, 4, 5])
b = np.array([6, 3, 4, 8, 9, 7, 1])

print(pair_max(a,b))
```

OUTPUT:

```
[6 7 9 8 9 7 5]
```

16. How to swap two columns in a 2d numpy array?

CODE:

```python
import numpy as np

arr = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9]).reshape(3,3)

print('Original array')
print(arr)

print("\nModified array")
print(arr[:, [1,0,2]])
```

OUTPUT:

```
Original array
[[1 2 3]
 [4 5 6]
 [7 8 9]]

Modified array
[[2 1 3]
 [5 4 6]
 [8 7 9]]
```

17. How to swap two rows in a 2d numpy array?

CODE:

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9]).reshape(3,3)

print('Original array')
print(arr)

print("\nModified array")
print(arr[[1,0,2], :])
```

OUTPUT:

```
Original array
[[1 2 3]
 [4 5 6]
 [7 8 9]]

Modified array
[[4 5 6]
 [1 2 3]
 [7 8 9]]
```

18. How to reverse the rows of a 2D array?

CODE:

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9]).reshape(3,3)

print('Original array')
print(arr)

print("\nModified array")
print(arr[::-1, :])
```

OUTPUT:

```
Original array
[[1 2 3]
 [4 5 6]
 [7 8 9]]

Modified array
[[7 8 9]
 [4 5 6]
 [1 2 3]]
```

19. How to reverse the columns of a 2D array?

CODE:

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9]).reshape(3,3)

print('Original array')
print(arr)

print("\nModified array")
print(arr[:, ::-1])
```

OUTPUT:

```
Original array
[[1 2 3]
 [4 5 6]
 [7 8 9]]

Modified array
[[3 2 1]
 [6 5 4]
 [9 8 7]]
```

20. How to create a 2D array containing random floats between 5 and 10?

CODE:

```
import numpy as np

rand_arr = np.random.uniform(5,10, size=(5,3))
print(rand_arr)
```

OUTPUT:

```
[[9.9174674  5.15947961 5.63776077]
 [8.44054145 5.66035747 9.5789622 ]
 [7.82382003 9.33949773 9.24700596]
 [9.78151729 5.79745772 7.19646827]
 [6.29604342 5.08023313 5.33084151]]
```

21. How to print only 3 decimal places in python numpy array?

CODE:

```
import numpy as np

rand_arr = np.random.random((3,2))
np.set_printoptions(precision=3)
print(rand_arr)
```

OUTPUT:

```
[[0.422 0.923]
 [0.888 0.235]
 [0.513 0.844]]
```

22. How to pretty print a numpy array by suppressing the scientific notation (like 1e10)?

CODE:

```
import numpy as np

np.random.seed(100)
rand_arr = np.random.random([3,3])/1e3
b=np.set_printoptions(suppress=True)
print(rand_arr)
```

OUTPUT:

```
[[0.0005434  0.00027837 0.00042452]
 [0.00084478 0.00000472 0.00012157]
 [0.00067075 0.00082585 0.00013671]]
```

23. How to limit the number of items printed in output of numpy array?

CODE:

```
import numpy as np

a = np.arange(15)
np.set_printoptions(threshold=6)
print(a)
```

OUTPUT:

```
[ 0  1  2 ... 12 13 14]
PS C:\Users\Raghavi\Documents>
```

24. How to print the full numpy array without truncating?

CODE:

```
import numpy as np

a = np.arange(15)
print(a)
```

OUTPUT:

```
[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14]
```

25. How to import a dataset with numbers and texts keeping the text intact in python numpy?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')
names = ('sepallength', 'sepalwidth', 'petallength', 'petalwidth', 'species')

# Print the first 3 rows
print(iris[:3])
```

OUTPUT:

```
[[b'5.1' b'3.5' b'1.4' b'0.2' b'Iris-setosa']
 [b'4.9' b'3.0' b'1.4' b'0.2' b'Iris-setosa']
 [b'4.7' b'3.2' b'1.3' b'0.2' b'Iris-setosa']]
```

26. How to extract a particular column from 1D array of tuples?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_1d = np.genfromtxt(url, delimiter=',', dtype=None)
species = np.array([row[4] for row in iris_1d])
print(species[:5])
```

OUTPUT:

```
Reading unicode strings without specifying the encoding argument i
s deprecated. Set the encoding, use None for the system default.
  iris_1d = np.genfromtxt(url, delimiter=',', dtype=None)
[b'Iris-setosa' b'Iris-setosa' b'Iris-setosa' b'Iris-setosa'
 b'Iris-setosa']
```

27. How to convert a 1d array of tuples to a 2d numpy array?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_1d = np.genfromtxt(url, delimiter=',', dtype=None)
iris_2d = np.array([row.tolist()[:4] for row in iris_1d])
print(iris_2d[:4])
```

OUTPUT:

```
[[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]]
```

28. How to compute the mean, median, standard deviation of a numpy array?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')
sepallength = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0])

print(np.mean(sepallength))
print(np.median(sepallength))
print(np.std(sepallength))
```

OUTPUT:

```
5.843333333333334
5.8
0.8253012917851409
```

29. How to normalize an array so the values range exactly between 0 and 1?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
sepallength = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0])

Smax, Smin = sepallength.max(), sepallength.min()
print((sepallength - Smin)/(Smax - Smin))
```

OUTPUT:

```
[0.22222222 0.16666667 0.11111111 0.08333333 0.19444444 0.30555556
 0.08333333 0.19444444 0.02777778 0.16666667 0.30555556 0.13888889
 0.13888889 0.         0.41666667 0.38888889 0.30555556 0.22222222
 0.38888889 0.22222222 0.30555556 0.22222222 0.08333333 0.22222222
 0.13888889 0.19444444 0.19444444 0.25       0.25       0.11111111
 0.13888889 0.30555556 0.25       0.33333333 0.16666667 0.19444444
 0.33333333 0.16666667 0.02777778 0.22222222 0.19444444 0.05555556
 0.02777778 0.19444444 0.22222222 0.13888889 0.22222222 0.08333333
 0.27777778 0.19444444 0.75       0.58333333 0.72222222 0.33333333
 0.61111111 0.38888889 0.55555556 0.16666667 0.63888889 0.25
 0.19444444 0.44444444 0.47222222 0.5        0.36111111 0.66666667
 0.36111111 0.41666667 0.52777778 0.36111111 0.44444444 0.5
 0.55555556 0.5        0.58333333 0.63888889 0.69444444 0.66666667
 0.47222222 0.38888889 0.33333333 0.33333333 0.41666667 0.47222222
 0.30555556 0.47222222 0.66666667 0.55555556 0.36111111 0.33333333
 0.33333333 0.5        0.41666667 0.19444444 0.36111111 0.38888889
 0.38888889 0.52777778 0.22222222 0.38888889 0.55555556 0.41666667
 0.77777778 0.55555556 0.61111111 0.91666667 0.16666667 0.83333333
 0.66666667 0.80555556 0.61111111 0.58333333 0.69444444 0.38888889
 0.41666667 0.58333333 0.61111111 0.94444444 0.94444444 0.47222222
 0.72222222 0.36111111 0.94444444 0.55555556 0.66666667 0.80555556
 0.52777778 0.5        0.58333333 0.80555556 0.86111111 1.
 0.58333333 0.55555556 0.5        0.94444444 0.55555556 0.58333333
 0.47222222 0.72222222 0.66666667 0.72222222 0.41666667 0.69444444
 0.66666667 0.66666667 0.55555556 0.61111111 0.52777778 0.44444444]
```

30. How to compute the softmax score?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
sepallength = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0])

def softmax(x):
    """Compute softmax values for each sets of scores in x.
    https://stackoverflow.com/questions/34968722/how-to-implement-the-softmax-function-in-python"""
    e_x = np.exp(x - np.max(x))
    return e_x / e_x.sum(axis=0)

print(sum(softmax(sepallength)))
```

OUTPUT:

```
0.9999999999999999
```

31. How to find the percentile scores of a numpy array?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
sepallength = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0])

print(np.percentile(sepallength, q=[5, 95]))
```

OUTPUT:

```
[4.6   7.255]
```

32. How to insert values at random positions in an array?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='object')

for i in np.random.randint(0, len(iris_2d), 20):
    iris_2d[i]=np.nan
print(iris_2d)
```

OUTPUT:

```
[[b'5.1' b'3.5' b'1.4' b'0.2' b'Iris-setosa']
 [b'4.9' b'3.0' b'1.4' b'0.2' b'Iris-setosa']
 [b'4.7' b'3.2' b'1.3' b'0.2' b'Iris-setosa']
 [nan nan nan nan nan]
 [nan nan nan nan nan]
 [b'5.4' b'3.9' b'1.7' b'0.4' b'Iris-setosa']
```

.
.
.

```
 [b'6.5' b'3.0' b'5.2' b'2.0' b'Iris-virginica']
 [b'6.2' b'3.4' b'5.4' b'2.3' b'Iris-virginica']
 [b'5.9' b'3.0' b'5.1' b'1.8' b'Iris-virginica']]
```

33. How to find the position of missing values in numpy array?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='float')
iris_2d[np.random.randint(150, size=20), np.random.randint(4, size=20)] = np.nan

np.isnan(iris_2d[:, 0]).sum()
print(np.where(np.isnan(iris_2d[:, 0])))
```

OUTPUT:

```
(array([  6,  53,  76, 107], dtype=int64),)
```

34. How to filter a numpy array based on two or more conditions?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0,1,2,3])

condition = (iris_2d[:, 2] > 1.5) & (iris_2d[:, 0] < 5.0)
print(iris_2d[condition])
```

OUTPUT:

```
[[4.8 3.4 1.6 0.2]
 [4.8 3.4 1.9 0.2]
 [4.7 3.2 1.6 0.2]
 [4.8 3.1 1.6 0.2]
 [4.9 2.4 3.3 1. ]
 [4.9 2.5 4.5 1.7]]
```

35. How to drop rows that contain a missing value from a numpy array?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0,1,2,3])

print(iris_2d[np.sum(np.isnan(iris_2d), axis = 1) == 0][:5])
```

OUTPUT:

```
[[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.  3.6 1.4 0.2]]
```

36. How to find the correlation between two columns of a numpy array?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris_2d = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0,1,2,3])
iris_2d[:,[0]]

np.corrcoef(iris_2d[:, 0], iris_2d[:, 2])
corr = np.corrcoef(iris_2d[:, 0], iris_2d[:, 2])[0, 1]
print(corr)
```

OUTPUT:

```
0.8717541573048712
```

37. How to get the second largest value of an array when grouped by another array?

CODE:

```
import numpy as np

nums = np.random.rand(5, 5)
print(nums)
second_largest = np.partition(nums, -2, axis=1)[:, -2]
print("\nSecond-largest value in each row:")
print(second_largest)
```

OUTPUT:

```
[[0.66142225 0.48116058 0.1151937  0.01554442 0.83145983]
 [0.07976305 0.1246498  0.33570137 0.91493868 0.75937255]
 [0.17758227 0.0451442  0.17881091 0.06486055 0.37099504]
 [0.12074221 0.25350194 0.35182247 0.03937923 0.2305162 ]
 [0.22622998 0.36620797 0.12361491 0.91454687 0.45704655]]

Second-largest value in each row:
[0.66142225 0.75937255 0.17881091 0.25350194 0.45704655]
```

38. How to sort a 2D array by a column?

CODE:

```
import numpy as np

arr = np.array([[100, 101, 500, 104],
                [201, 202, 203, 204],
                [301, 300, 600, 307]])
Index = 2
Array_sort = arr[arr[:,Index].argsort()]
print("The sorted array is:", "\n", "\n", Array_sort)
```

OUTPUT:

```
The sorted array is:

 [[201 202 203 204]
 [100 101 500 104]
 [301 300 600 307]]
```

39. How to find the most frequent value in a numpy array?

CODE:

```
import numpy as np

x = np.array([1,2,3,4,5,1,2,1,1,1])
print("Original array:")
print(x)

print("Most frequent value in the above array:")
print(np.bincount(x).argmax())
```

OUTPUT:

```
Original array:
[1 2 3 4 5 1 2 1 1 1]
Most frequent value in the above array:
1
```

40. How to find the position of the first occurrence of a value greater than a given value?

CODE:

```
import numpy as np

url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='object')

res = np.where(iris[:,3].astype(float) > 1)[0][0]
print(res)
```

OUTPUT:

```
50
```

41. How to replace all values greater than a given value to a given cutoff?

CODE:

```
import numpy as np

x = np.array([[ 0.42436315, 0.48558583, 0.32924763],
              [ 0.7439979,0.58220701,0.38213418],
              [ 0.5097581,0.34528799,0.1563123 ]])
print("Original array:")
print(x)
print("Replace all elements of the said array with .5 which are greater than .5")
x[x > .5] = .5
print(x)
```

OUTPUT:

```
Original array:
[[0.42436315 0.48558583 0.32924763]
 [0.7439979  0.58220701 0.38213418]
 [0.5097581  0.34528799 0.1563123 ]]
Replace all elements of the said array with .5 which are greater than .5
[[0.42436315 0.48558583 0.32924763]
 [0.5        0.5        0.38213418]
 [0.5        0.34528799 0.1563123 ]]
```

42. How to get the positions of top n values from a numpy array?

CODE:

```
import numpy as np

arr = np.array([1, 3, 2, 4, 5])
print(arr.argsort()[-3:][::-1])
```

OUTPUT:

```
[4 3 1]
```

43. How to compute the row wise counts of all possible values in an array?

CODE:

```
import numpy as np

np.random.seed(100)
arr = np.random.randint(1,11,size=(6, 10))

row,col = np.shape(arr)
for i in range(row):
    c = np.bincount(arr[i], minlength=11)[1:]
    print(c)
```

OUTPUT:

```
[1 0 2 1 1 1 0 2 2 0]
[2 1 3 0 1 0 1 0 1 1]
[0 3 0 2 3 1 0 1 0 0]
[1 0 2 1 0 1 0 2 1 2]
[2 2 2 0 0 1 1 1 1 0]
[1 1 1 1 1 2 0 0 2 1]
```

44. How to convert an array of arrays into a flat 1d array?

CODE:

```
import numpy as np

arr1 = np.arange(3)
arr2 = np.arange(3,7)
arr3 = np.arange(7,10)

arr_2d = np.concatenate([arr1, arr2, arr3])
print(arr_2d)
```

OUTPUT:

```
[0 1 2 3 4 5 6 7 8 9]
```