

PDAT 617 Assignment 2 Due on 06/22 11:59pm

1. Finish the following questions in numpy. (Hint: you can use function directly, like `numpy.full()` to get full array.) (50 points)

- **1):** Write a NumPy program to create an empty and a full array. (10 points)
- **2):** Write a NumPy program to get the unique elements of an array. (10 points).
For example: Original array: [10 10 20 20 30 30]
Unique elements of the above array: [10 20 30]
- **3):** Write a NumPy program to construct an array by repeating. (10 points)
Sample array: [1, 2, 3, 4]
Expected Output: Repeating 2 times [1 2 3 4 1 2 3 4]
Repeating 3 times [1 2 3 4 1 2 3 4 1 2 3 4]
- **4):** Write a NumPy program to sort pairs of first name and last name return their indices. (first by last name, then by first name). (10 points)
firstnames = (Betsey, Shelley, Lanell, Genesis, Margery)
lastnames = (Battle, Brien, Plotner, Stahl, Woolum)
Expected Output: [1 3 2 4 0]
- **5):** Write a Python program to count number of occurrences of each value in a given array of non-negative integers. (10 points)
Original array: [0 1 6 1 4 1 2 2 7]
Number of occurrences of each value in array:
1 3 2 0 1 0 1 1

2. Answer the following question with iris dataset. (50 points)

Copy the following code to read dataset:

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

- 1). Find the mean, median, standard deviation of iris's sepalength (1st column). (10 points)
- 2). Find the 5th and 95th percentile of iris's sepalength. (10 points)
- 3). Create a normalized form of iris's sepalength whose values range exactly between 0 and 1 so that the minimum has value 0 and maximum has value 1. (10 points)
- 4). Filter the rows of iris that has petalength (3rd column) > 1.5 and sepalength (1st column) < 5.0. (20 points)