

CS 795/895 –Introduction to Data Science, HW2

- A. (80 pts) Python programming. Please use the [code skeleton](#) provided
- (20 pts) Write a python function to merge 2 sorted lists. The resultant list should be in sorted order.
 - (30 pts) Write a python function to calculate the summary statistics of an array of numbers: max, min, mean, standard deviation, median, 75 percentiles, 25 percentiles. You can write your own functions or use any existing modules from math, statistics or numpy.
 - (30 pts) Re-scaling is the process of linearly transform from one range to another. Write a python function that accepts a list of numbers in any range, then scales the numbers to [0, 1] using MinMax algorithm (given in the skeleton code).
- B. (20 pts) Pandas basics

Let df be a pandas DataFrame constructed with the following code:

```
data = np.array([0, 7, 3, 6, 2, 8, 5, 9, 4]).reshape(3, 3)
```

```
df = pd.DataFrame(data, index=['One', 'Two', 'Three'], columns=['a', 'b', 'c'])
```

What is the output of the following code? (Try to write the output without using python first)

- print(df)
- df['a']
- df.loc['Two']
- df[:2]
- df.iloc[:,:]
- df.iloc[:,2]
- list(df.columns)
- list(df.index)
- df['b']['Two']
- list(df.iloc[2, :])
- df.drop('a', axis=1)
- df[df.a != 5]
- list(df.sum(axis=0))
- df.iloc[:, list(df.sum(axis=0) < 17)]
- df.sort_values(by='c')
- df.sort_values(by='Two', axis=1)
- df.T
- (df<=2).any(axis=0)
- df.applymap(lambda x: x*2-1)
- df.apply(lambda x: max(x), axis=1)

What to turn in:

Each file must exactly follow the naming convention: **Lastname-hw2.zip** should contain following 2 files.

- HW2a.py
- HW2b.pdf