

## CSYE7105 HW3

Instructor: Dr. Handan Liu

2020-11-1

The following parts are required to be written in python file and run in parallel mode on the Discovery Cluster.

**Part 1: 30 points**

1. Use Pool.apply() to get the row wise common items in list\_a and list\_b; and print the result. (10 pts)

```
list_a = [[1, 2, 3], [5, 6, 7, 8], [10, 11, 12], [20, 21]]
```

```
list_b = [[2, 3, 4, 5], [6, 9, 10], [11, 12, 13, 14], [21, 24, 25]]
```

2. Use Pool.map() to run the following python scripts in parallel; and print the result. (10 pts)

Script names: 'script1.py', 'script2.py', 'script3.py'

Hint: you can put any content in the three scripts.

3. Normalize each row of 2d array (list) list\_c to vary between 0 and 1. Parallelize the function with any subfunction of Pool; and print the result. (10 pts)

```
list_c = [[2, 3, 4, 5], [6, 9, 10, 12], [11, 12, 13, 14], [21, 24, 25, 26]]
```

**Part 2: 20 points**

A serial code is given here: HW3-p2.py. Please complete the following tasks:

- Parallelize it with Pool.apply when CPU=4, and save as HW3-p2\_apply.py (10 pts)
- Parallelize it with Pool.startmap when CPU=4, and save as HW3-p2\_startmap.py (10 pts)

### Part 3: 20 points

A given panda dataframe here:

```
df = pd.DataFrame(np.random.randint(3, 10, size=[20000, 100]))
```

Please complete the following tasks:

- Define a function to sum the squares of each row, and then calculate the square root, leaving two decimal places for the result. (10 pts)
- Parallelize the function with Pool.imap, and print the elapsed time when CPU=2, 4 and 8. (10 pts)

### Part 4: 30 points

Parallel tuning C parameter in SVM:

Implement a naive parallel k-fold cross-validation algorithm in Python for tuning the "C" parameter in Support Vector Machines. The "kfold-cv-pool.py" is given. The dataset is "optdigits.txt".

- 1) Please use for-loop to set 'Pool' parallelism (see 'kfold-cv-pool.py') working on CPU=2,4,8. Calculate the elapsed time to plot that the elapsed time (y-axis) as the number of processors (x-axis) is varied, including 1 processor with the serial code. (10 pts)
- 2) Please change the parallelism using 'Process'. (20 pts)

You can put the 2 modes in one file; or in separately files.

### Review and Grade:

TA will review the homework including run the code on Discovery. TA will grade for my reference. No review time in person or on Zoom.

Submission format: compact all files as a tarball with your name for the convenience of TA

Submission through Canvas

Deadline is by the end of November 9<sup>th</sup>.