


CME 211 - 102

BECAUSE WE WANT TO GO FAST

“WHY IS MY PROGRAM SLOW?”

- BECAUSE OF THE WRONG DATA STRUCTURES
- BECAUSE OF IMBRICATED FOR LOOPS
- BECAUSE OF BAD CODE ARCHITECTURE
- ETC...

WHY CAN IT IMPORTANT TO HAVE FAST CODE

- TO ITERATE AND EXPERIMENT FASTER
- BECAUSE YOUR USER WILL BE ANGRY
- BECAUSE YOU LIKE SPEED 

HOW TO MAKE THINGS FASTER

- ~~• LOOK AT YOUR CODE FOR 3 HOURS TO TRY TO FOCUS HARD~~
- USE A PROFILER ← YES

IN PYTHON: USING CPROFILE AND PSTATS

python3 -m cProfile similarity.py u.data

Function called once but takes a long time

Function called a lot

```
Random listing order was used
ncalls  tottime  percall  cumtime  percall filename:lineno(function)
      2   0.032    0.016    0.032    0.016 {built-in method io.open}
      2   0.033    0.016    0.033    0.016 {method 'close' of '_io.TextIOWrapper' objects}
      1   0.000    0.000    0.000    0.000 codecs.py:259(__init__)
      1   0.456    0.456    0.568    0.568 similarity.py:3(read_data)
    1682   0.042    0.000    0.042    0.000 {method 'index' of 'list' objects}
      5   0.000    0.000    0.000    0.000 {built-in method builtins.print}
      1   0.000    0.000    41.262   41.262 {built-in method builtins.exec}
     243   0.003    0.000    0.003    0.000 {built-in method _codecs.utf_8_decode}
  200001   0.022    0.000    0.022    0.000 {method 'keys' of 'dict' objects}
      2   0.000    0.000    0.000    0.000 _bootlocale.py:23(getpreferredencoding)
  419789   1.089    0.000   13.750    0.000 similarity.py:48(corr_coef)
    1682   0.001    0.000    0.001    0.000 {method 'write' of '_io.TextIOWrapper' objects}
      2   0.000    0.000    0.000    0.000 {built-in method _locale.nl_langinfo}
  1259314   1.981    0.000   12.661    0.000 similarity.py:57(inner_prod)
  100000   0.025    0.000    0.025    0.000 {method 'add' of 'set' objects}
    1683   0.038    0.000    0.038    0.000 similarity.py:124(<genexpr>)
      2   0.000    0.000    0.000    0.000 {built-in method time.time}
      1   1.654    1.654   40.658   40.658 similarity.py:112(find_sim)
  1413721  10.362    0.000   38.656    0.000 similarity.py:82(calc_sim)
    1682   0.193    0.000    0.193    0.000 {built-in method builtins.max}
    1684   0.004    0.000    0.004    0.000 {method 'format' of 'str' objects}
      1   0.000    0.000    0.000    0.000 {method 'items' of 'dict' objects}
      1   0.000    0.000    0.000    0.000 {built-in method builtins.sorted}
  18134196   1.836    0.000    1.836    0.000 {method 'append' of 'list' objects}
     243   0.001    0.000    0.004    0.000 codecs.py:318(decode)
```


IN PYTHON: USING CPROFILE AND PSTATS

```
python3 -m cProfile -o profile_state similarity.py u.data
```

```
analyze_stats.py > ...
```

```
1 import pstats
2
3 p = pstats.Stats("profile_stats")
4 # stats.strip_dirs().sort_stats(-1).print_stats()
5 p.strip_dirs()
6 p.sort_stats('time')
7 p.print_callees()
8 # p.print_stats(10)
```

```
peytavin@rice09:/farmshare/user_data/peytavin/good_repo_hw2/hw2$ python3 analyze_stats.py
```

Ordered by: internal time

| Function | called... | ncalls | tottime | cumtime | |
|-----------------------------|-----------|----------|---------|---------|--|
| similarity.py:82(calc_sim) | -> | 419789 | 1.089 | 13.750 | similarity.py:48(corr_coef) |
| | | 839578 | 4.338 | 9.191 | similarity.py:107(<lambda>) |
| | | 1413721 | 0.122 | 0.122 | {built-in method builtins.len} |
| | | 18034196 | 1.821 | 1.821 | {method 'append' of 'list' objects} |
| | | 1413721 | 3.408 | 3.408 | {method 'intersection' of 'set' objects} |
| similarity.py:71(<genexpr>) | -> | | | | |

IN PYTHON: USING CPROFILE AND PSTATS

Solve the issue: Extract lambda function definition from within the loop

```
demean = lambda ls, mean: list(ls[i] - mean for i in range(len(ls)))
```

Rerun

| Function | called... | ncalls | tottime | cumtime | |
|-------------------------------|-----------|----------|---------|---------|--|
| similarity.py:84(calc_sim) | -> | 839578 | 4.253 | 9.032 | similarity.py:3(<lambda>) |
| | | 419789 | 1.064 | 13.843 | similarity.py:50(corr_coef) |
| | | 1413721 | 0.127 | 0.127 | {built-in method builtins.len} |
| | | 18034196 | 1.712 | 1.712 | {method 'append' of 'list' objects} |
| | | 1413721 | 3.288 | 3.288 | {method 'intersection' of 'set' objects} |
| similarity.py:73(<generator>) | -> | | | | |