# Directional Changes Project

## Follow up 30/07/2018

Jurgen Palsma

# What have I done since last time (03/07)?

- Implemented PSO
- Implemented CSFLA
- Scripted experiment
- KS2 test
- Started results analysis

# PSO implementation

- Implemented nearly everything described in early deliverable (except neighbourhood)

- Was very excited about results but realised it was on train data

- After param tweaking got good results in terms of fitness but very slow algorithm - not what I expected

# CSFLA implem

- Went pretty smoothly due to research done for the early deliverable and due to the project architecture (I could reuse PSO classes and focus on logic/algorithm)

- Very fast convergence but pretty mediocre results in terms of fitness

# Scripted Experiment

- Scripted everything so that I can configure through a json file:
  - GA parameters
  - PSO parameters
  - CSFLA parameters
  - Data files

- Even though the experiment takes a long time the configuration and scripting makes the process much less tedious

# Kolmogorov-Smirnov

- Looked at the KS test and found a library to execute the test on my data
- Not yet in experiment script

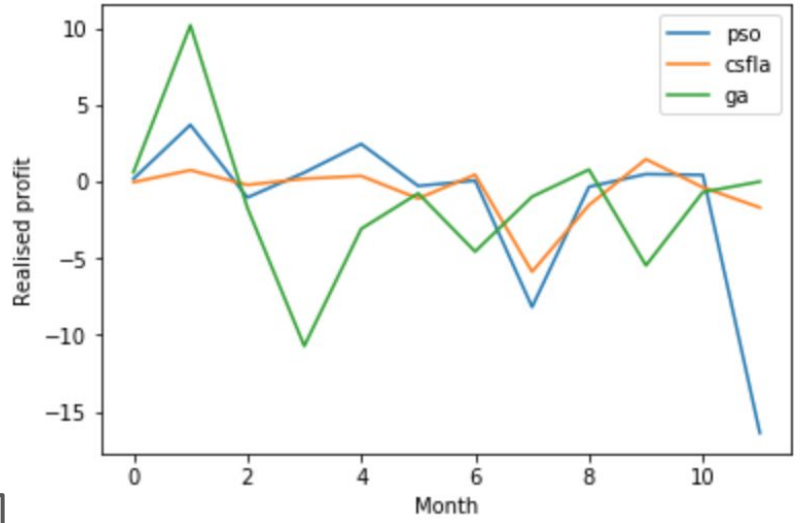
# Results analysis

- Looked into plotting libraries
- Started looking at results (following slides)
- NB: arbitrarily set experiment parameters

# Results analysis - returns

In terms of profit:

- Not counting the last month, PSO seems to perform better
- PSO and CSFLA seem correlated
- GA is sometimes seems inversely correlated to PSO/CSFLA




Realised Profit per Month for each algorithm

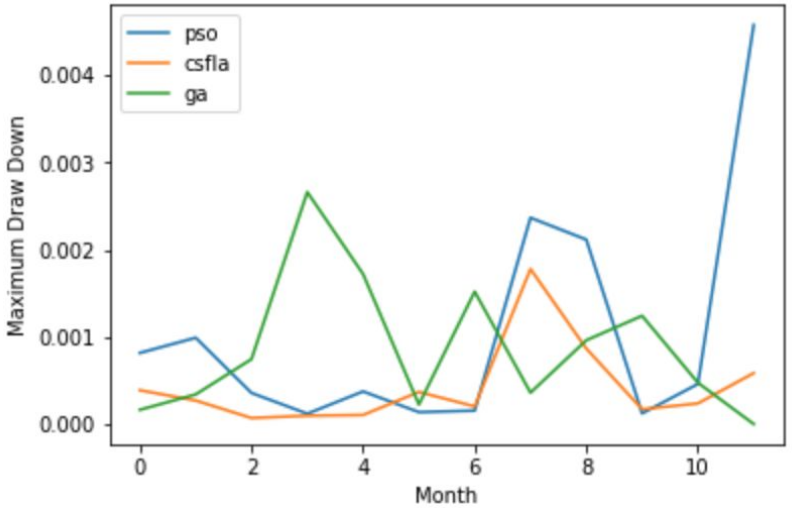
| algo | PSO | CSFLA | GA |
|---|---|---|---|
| Mean return | -0.17 | -0.53 | -1.48 |

Mean return on plotted results, not counting the last month

# Results analysis - MDD

In terms of MDD:

- PSO and CSFLA still seem correlated
- Some months are chaotic for each algorithm
- Even though PSO shoots at month 12, its relatively low MDD




| algo | PSO | CSFLA | GA |
|---|---|---|---|
| Mean MDD | 0.0010 | 0.0004 | 0.0009 |

Mean MDD on plotted results, for all months

# Questions - 1

- KS test:
  - What do I compare? The fitness distribution across each best individual for each run or all individuals in a/all run(s)

- What params should I set for the GA?
  - The same for each month (the set you provided per email)?
  - More/less training/testing days?

# Questions - 2

- How can I compare the algorithms?
  - My results depend on algorithm params, eg swarm size for PSO, but the higher the swarm size, the higher the computation needs.
  - Should I give the same execution time limit to each algorithm? Or something like same pop size?
  - Should I focus on given fitness results?

# What’s next (?)

- Is this enough? I believe I reached my objectives. Or should I:
  - Try to optimize my results (tune params)?
  - Optimize/clean up/finish my algorithms?
  - Start writing my dissertation?
  - Go more in depth with the results analysis?

- Otherwise, I could/would like to:
  - Try and set up the experiment in the university’s computer cluster
  - Set up DC on quantopian:
    - I’ve briefly started to try and code DCs
    - It gives access to a LOT of datasets and backtesting capacities
    - Allows to “**alpha-factor** test” DC