Assignment #2
Genetic Programming

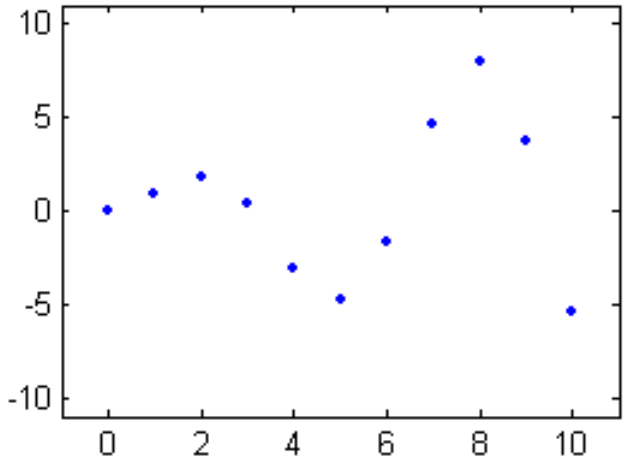
MECS 4510

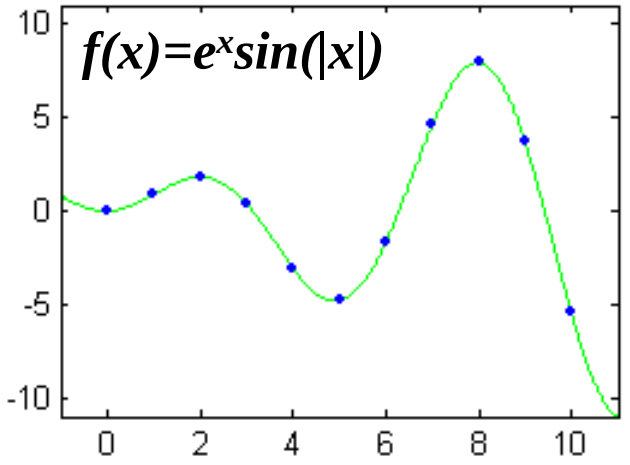
Evolutionary Computation

Hod Lipson

Symbolic Regression

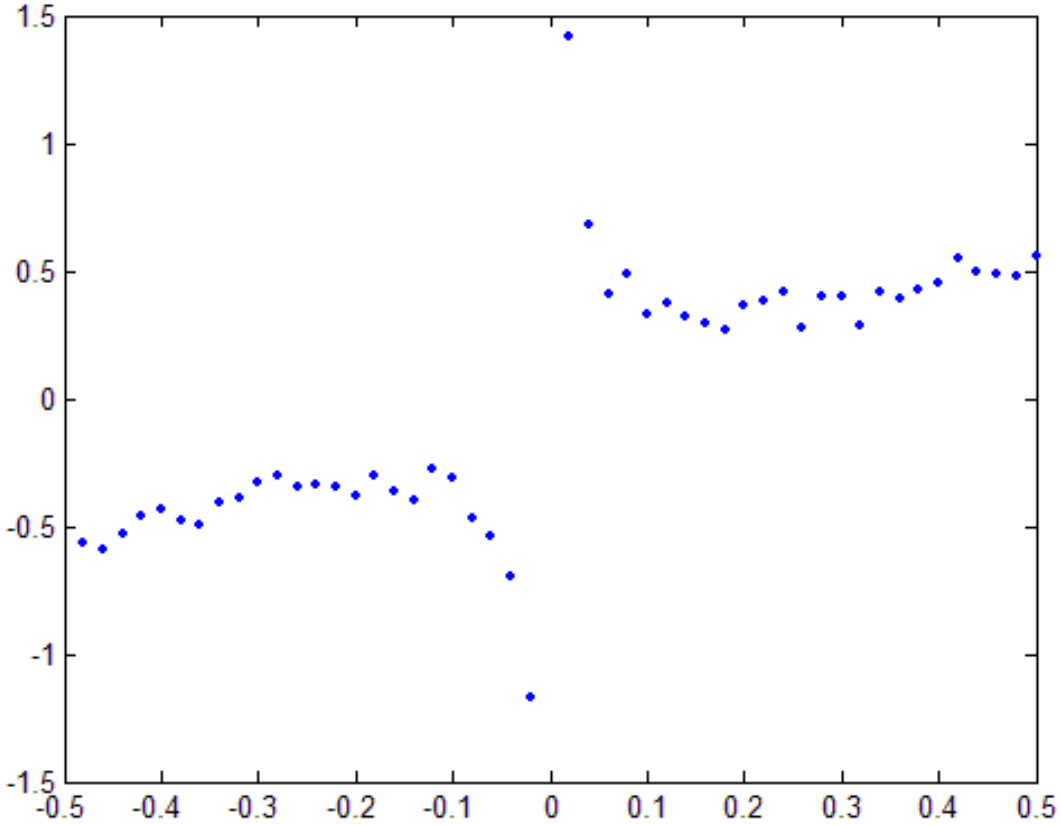
What function describes this data?



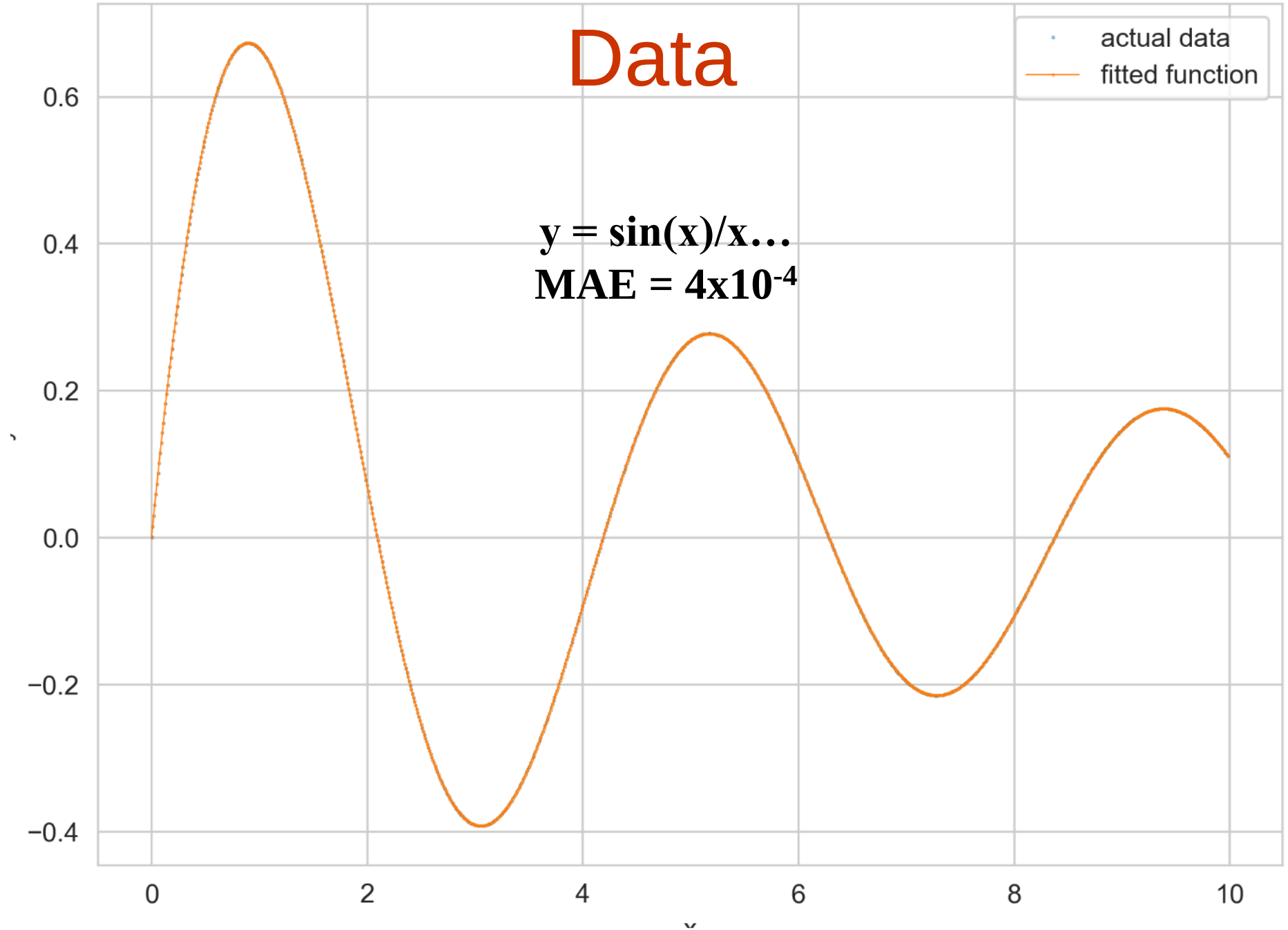


John Koza, 1992

Data



Data

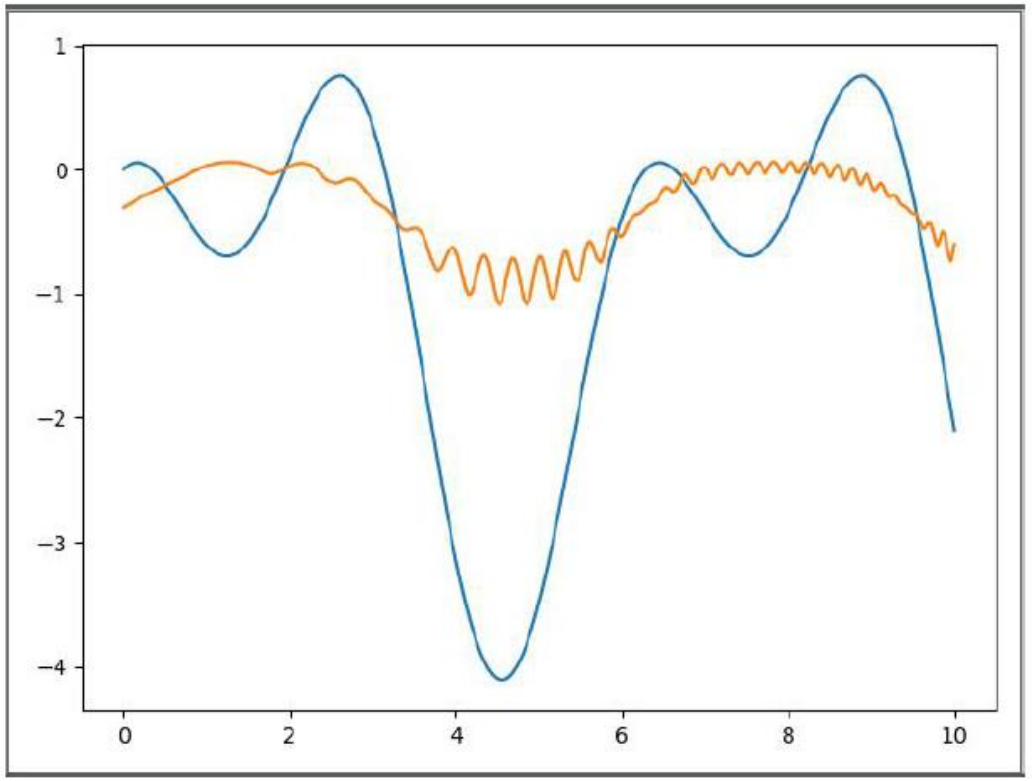


Boxi Xia

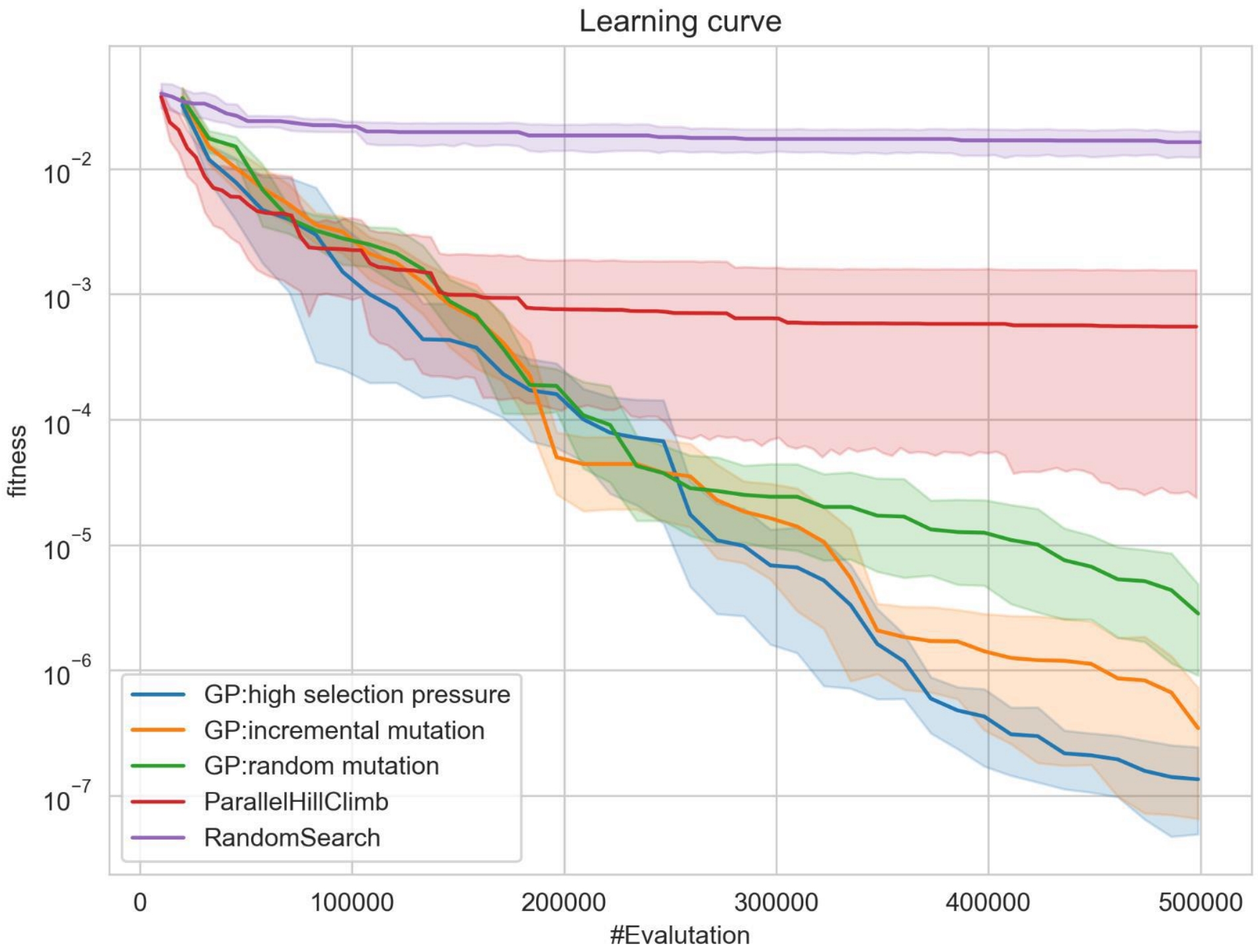
Week 1: Random function

equation = *cos(cos(cos(x*x))*sin(sin(x))+*
x/x(-9.247776198625752)+*
(-8.62238505749605)/(-4.15149813701656))-
sin((1.2185149095292367))

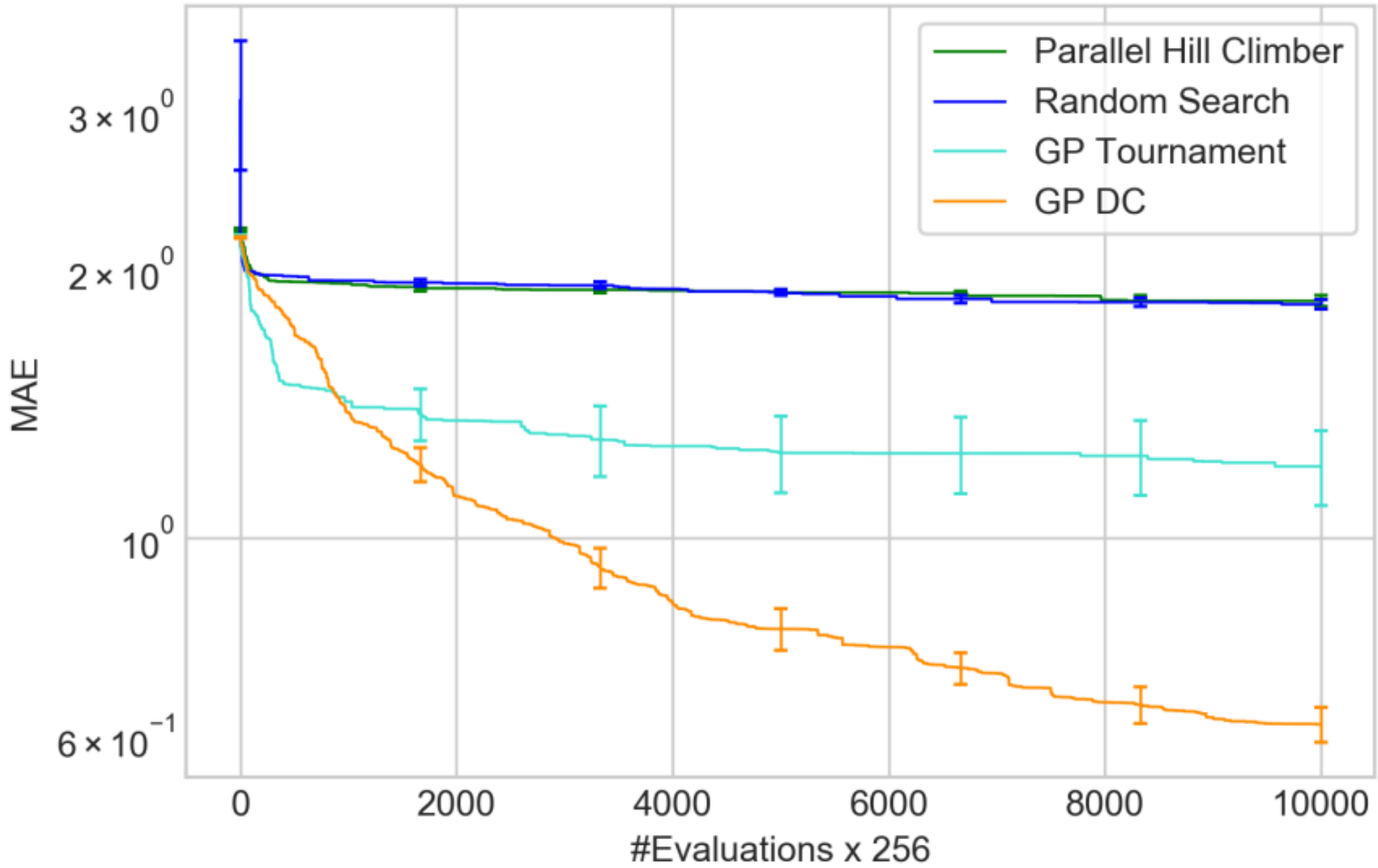
MSE = *1.5227438847355902*

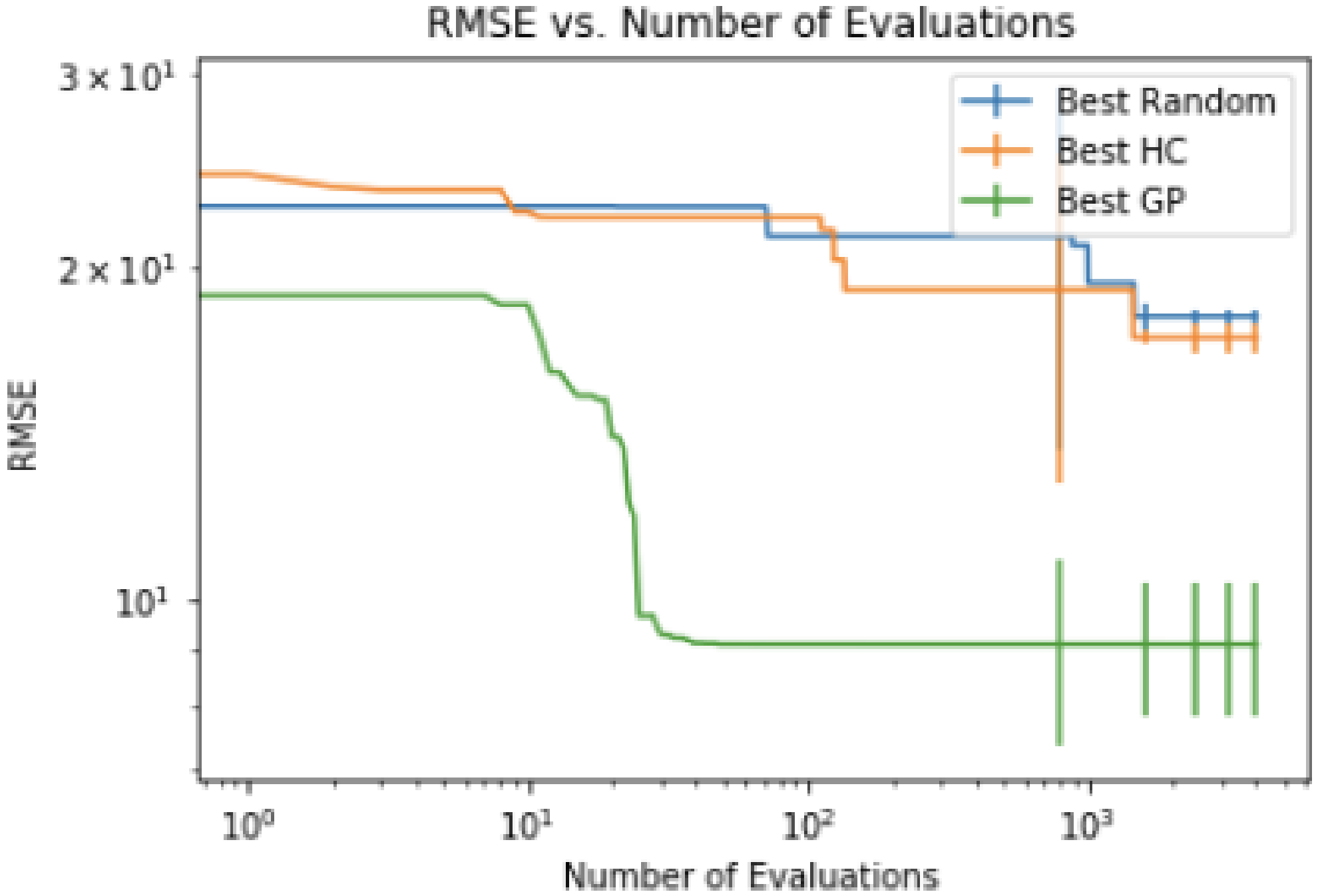


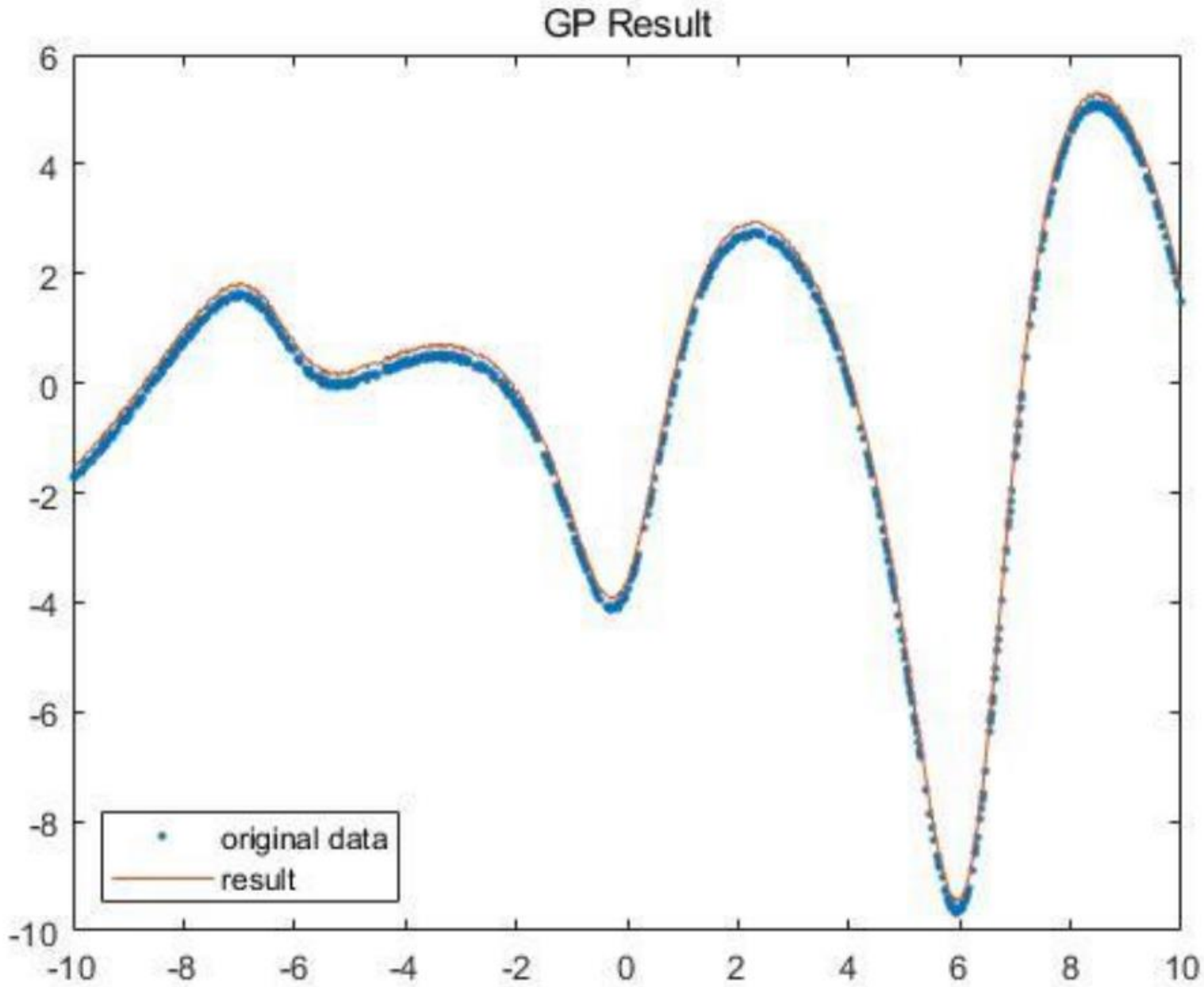
Seaborn

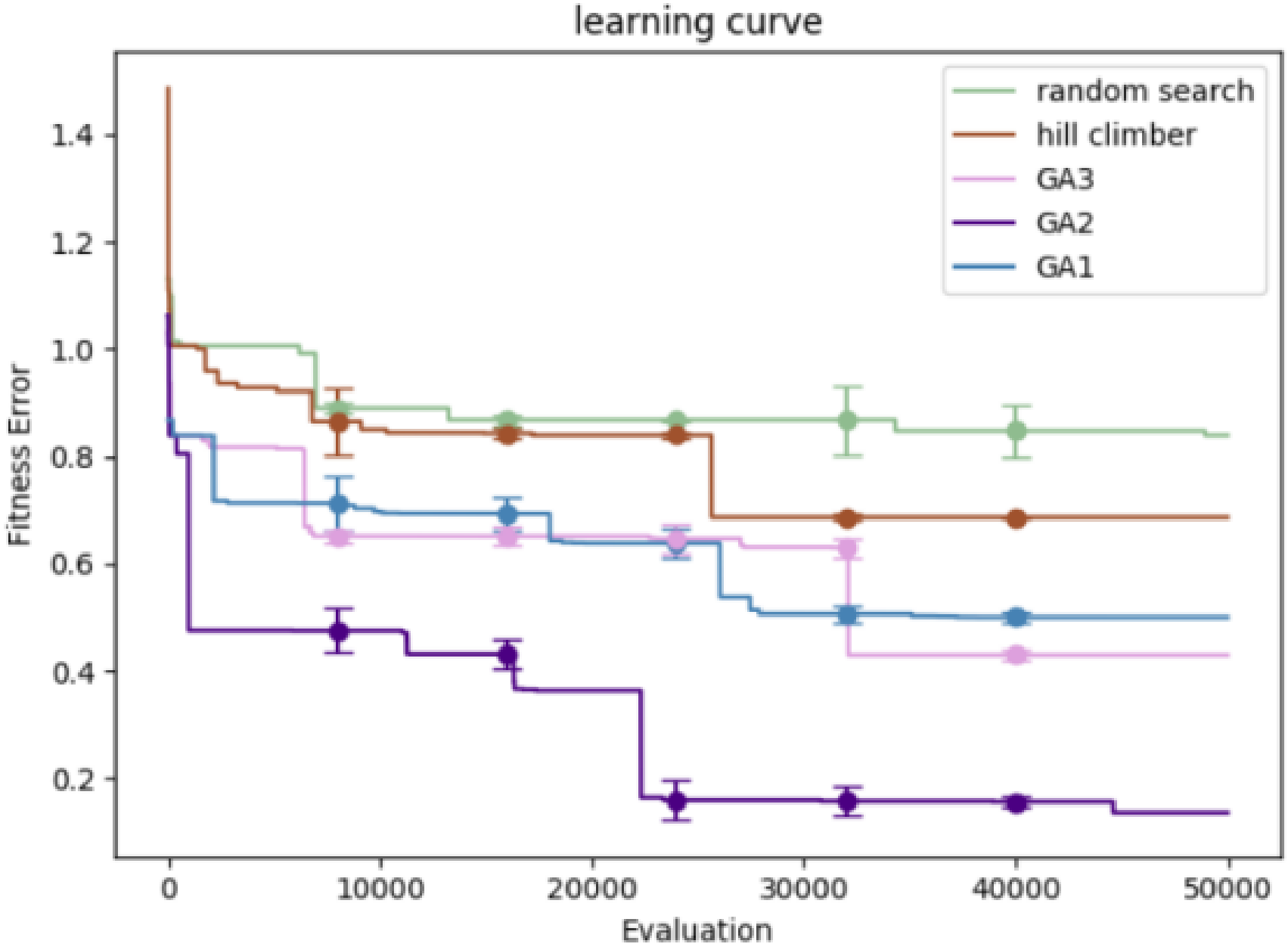


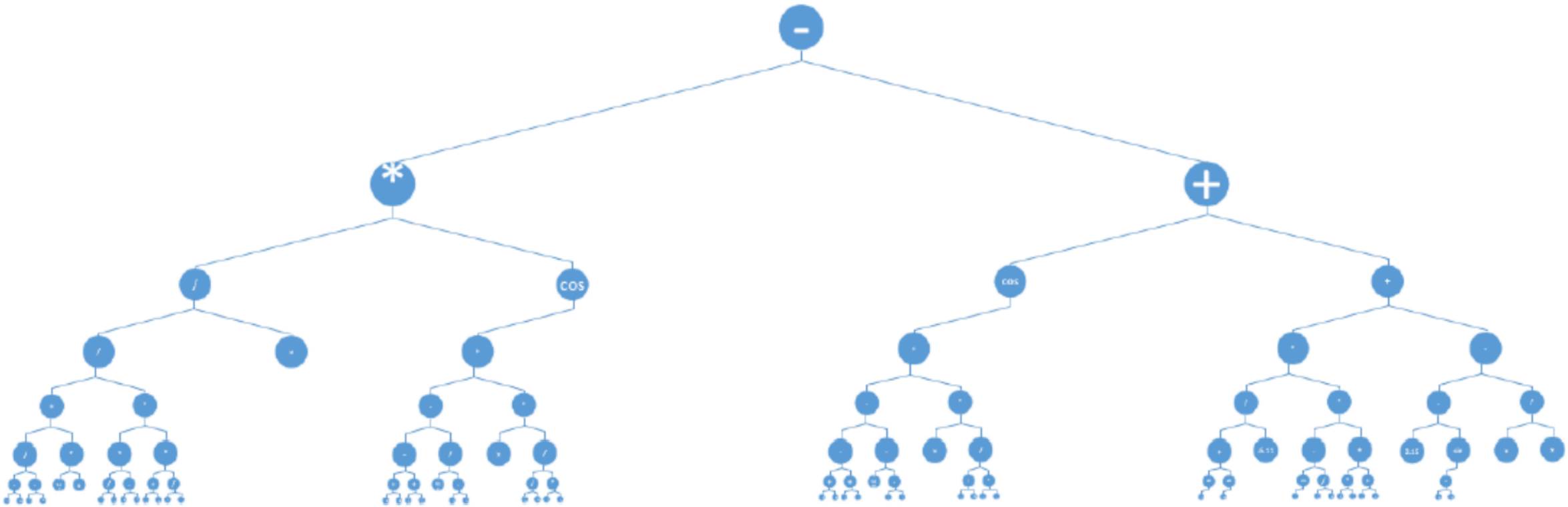
Boxi Xia

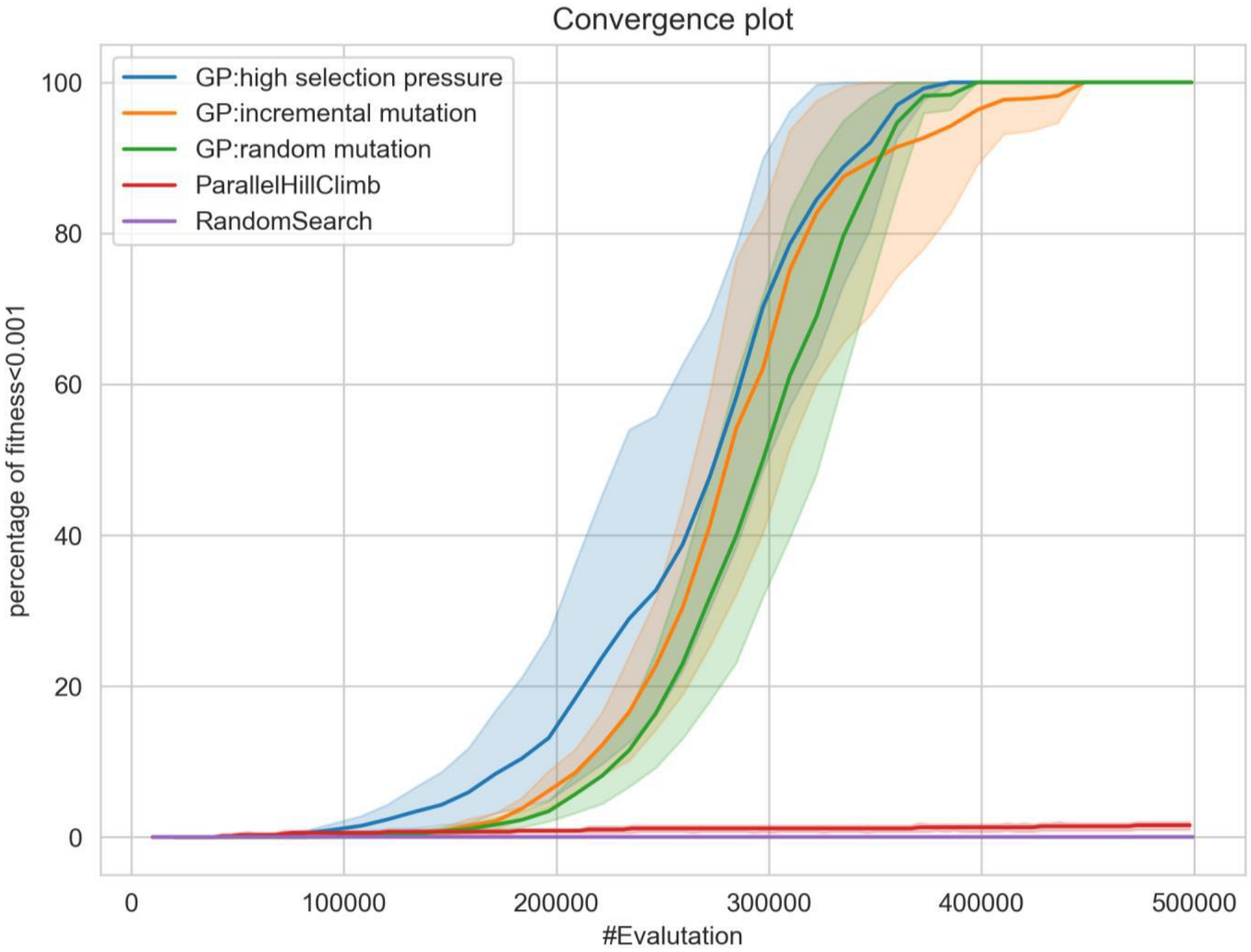

Parallel Hill Climber
Random Search
GP Tournament
GP DC
MAE
3×10^0
2×10^0
10^0
6×10^{-1}
0
2000
4000
6000
8000
10000
#Evaluations x 256


RMSE vs. Number of Evaluations
Best Random
Best HC
Best GP
RMSE
Number of Evaluations
3 × 10¹
2 × 10¹
10¹
10⁰
10¹
10²
10³

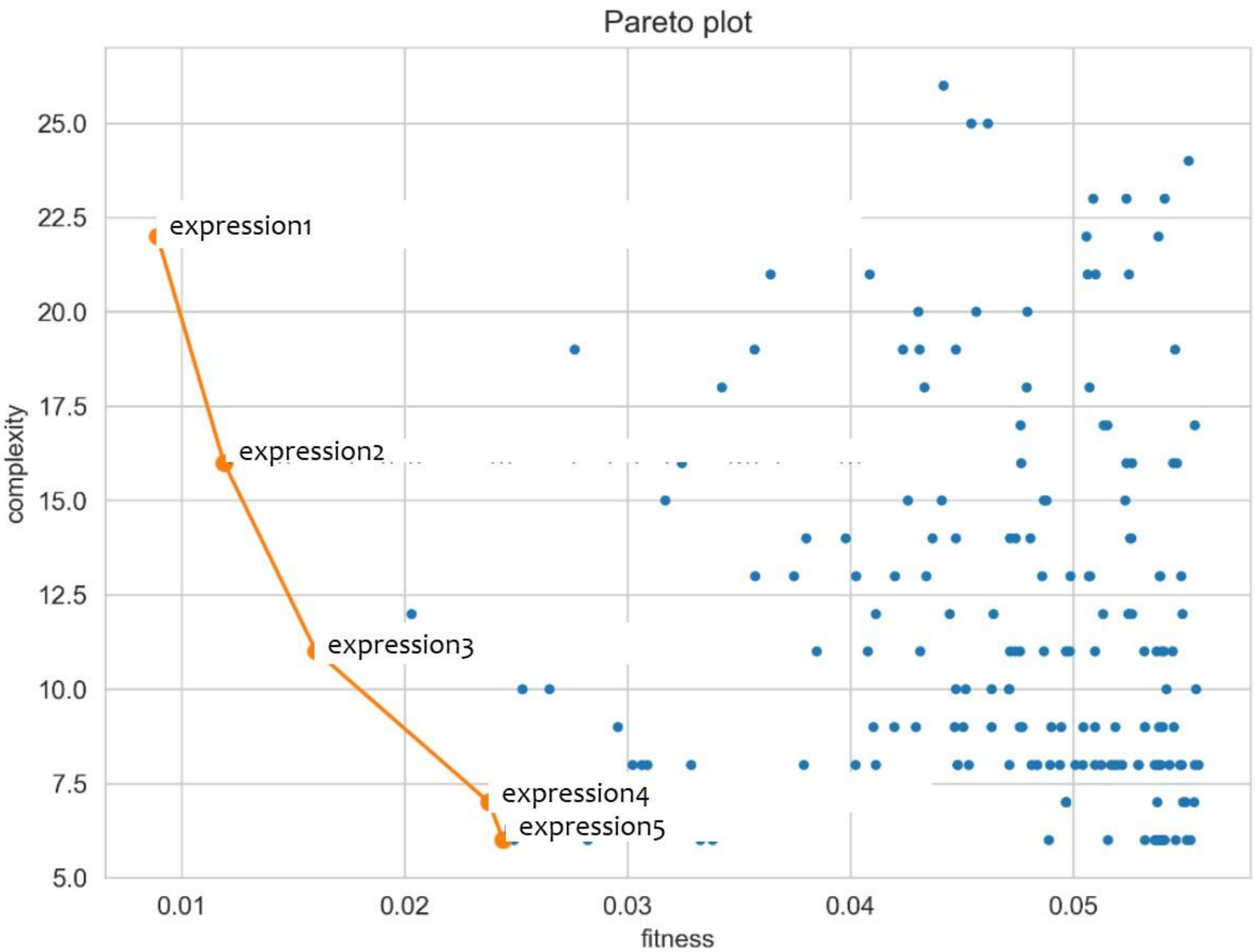

GP Result
6
4
2
0
-2
-4
-6
-8
-10
-10
-8
-6
-4
-2
0
2
4
6
8
10
original data
result


learning curve
random search
hill climber
GA3
GA2
GA1
Fitness Error
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
10000
20000
30000
40000
50000
Evaluation


−
*
+
/
cos
cos
+

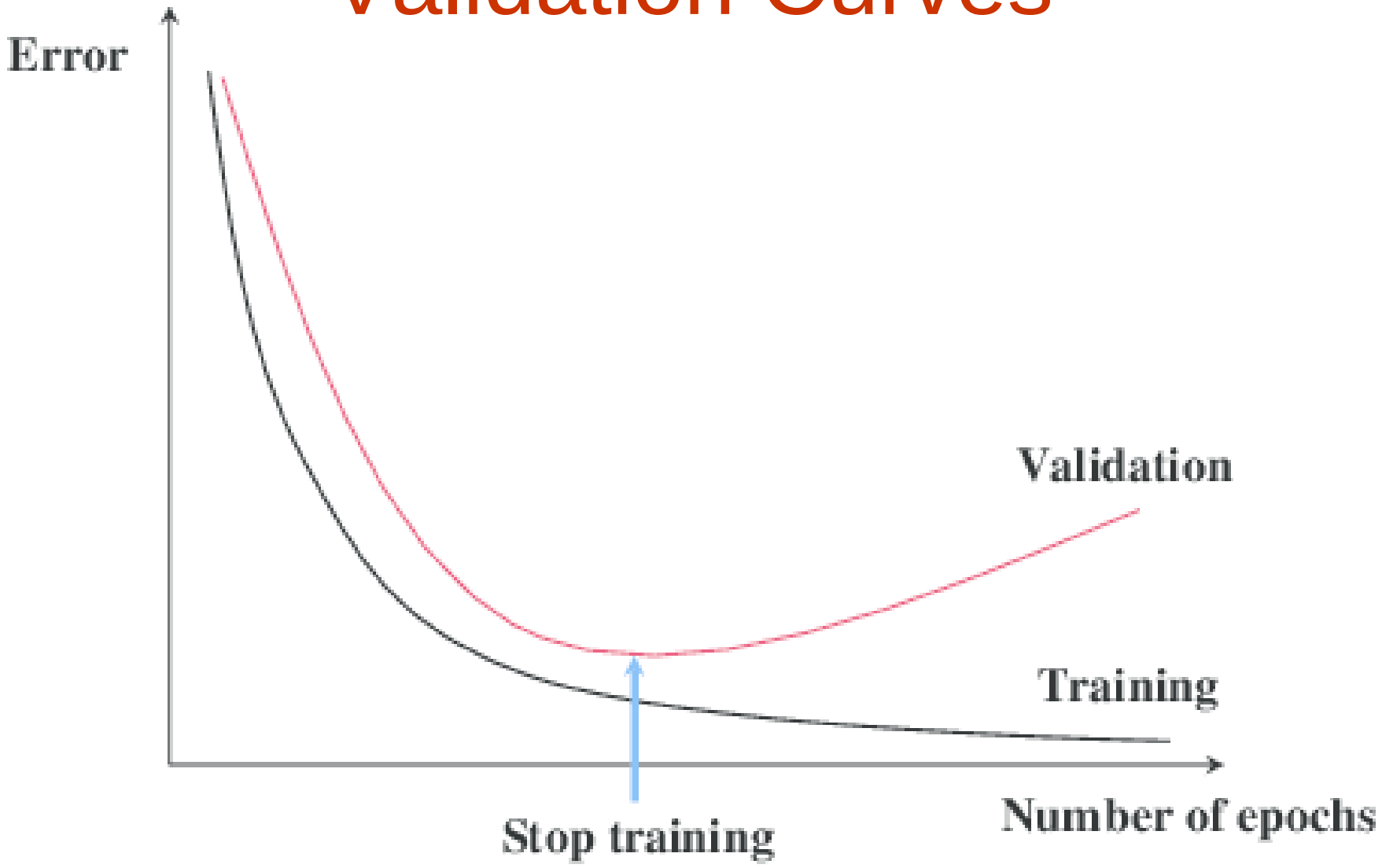

Convergence plot
GP:high selection pressure
GP:incremental mutation
GP:random mutation
ParallelHillClimb
RandomSearch
percentage of fitness<0.001
100
80
60
40
20
0
0
100000
200000
300000
400000
500000
#Evalutation

Boxi Xia

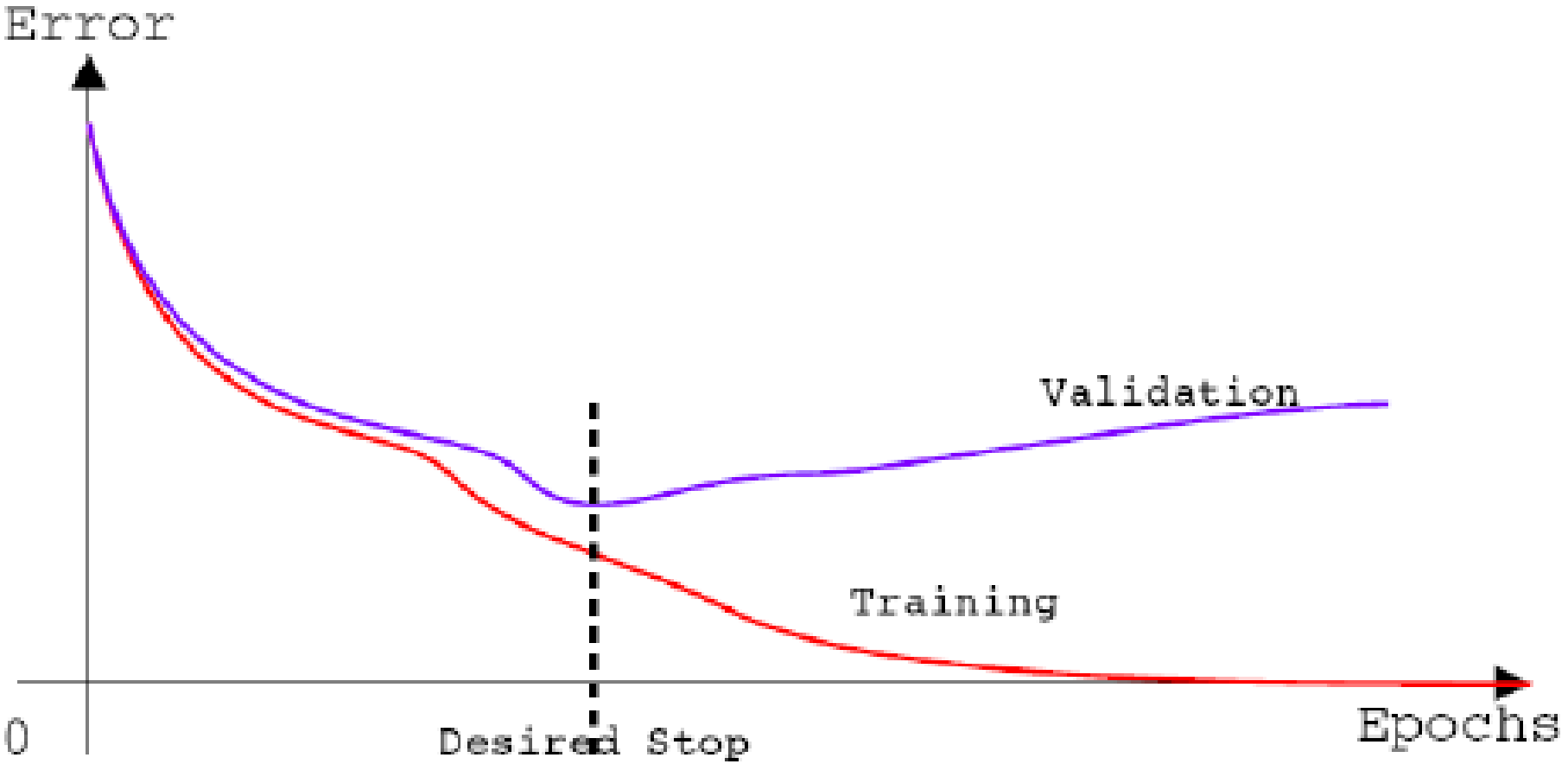

Pareto plot
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
complexity
expression1
expression2
expression3
expression4
expression5
0.01
0.02
0.03
0.04
0.05
fitness

Boxi Xia

Validation Curves

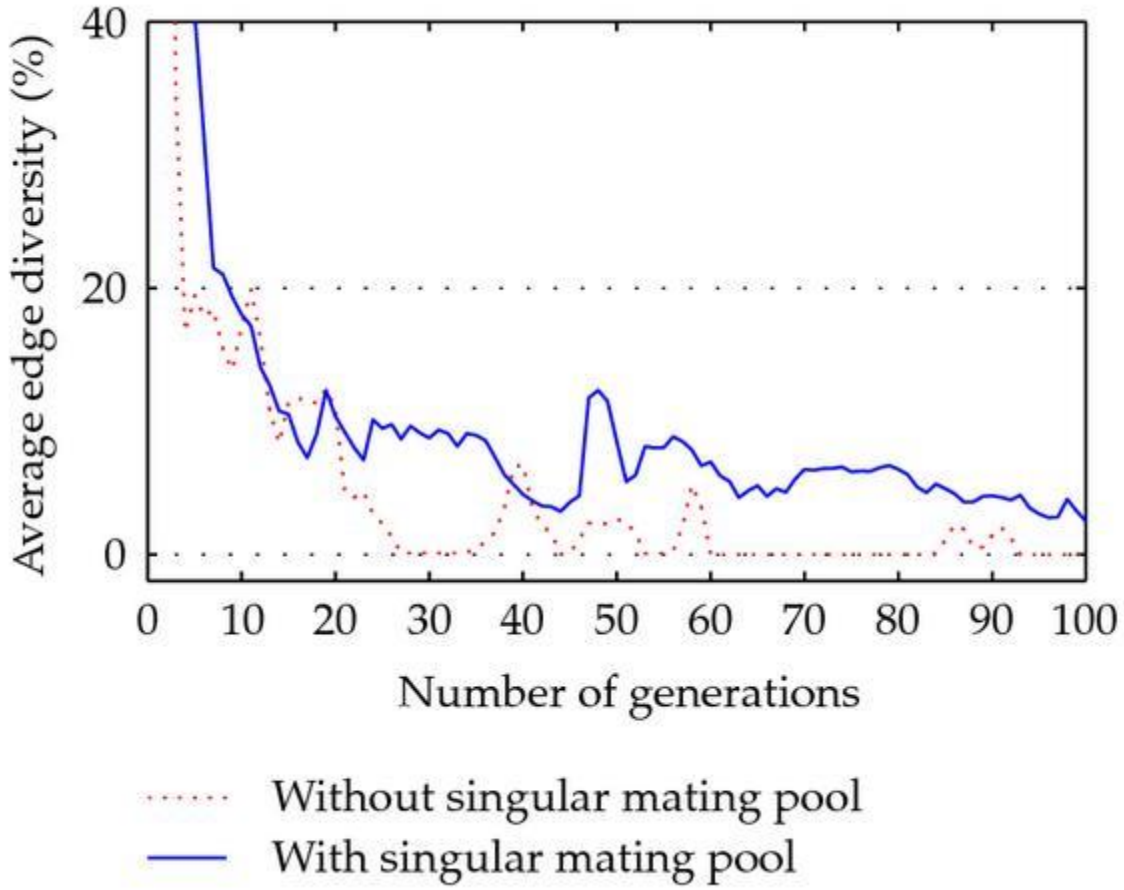


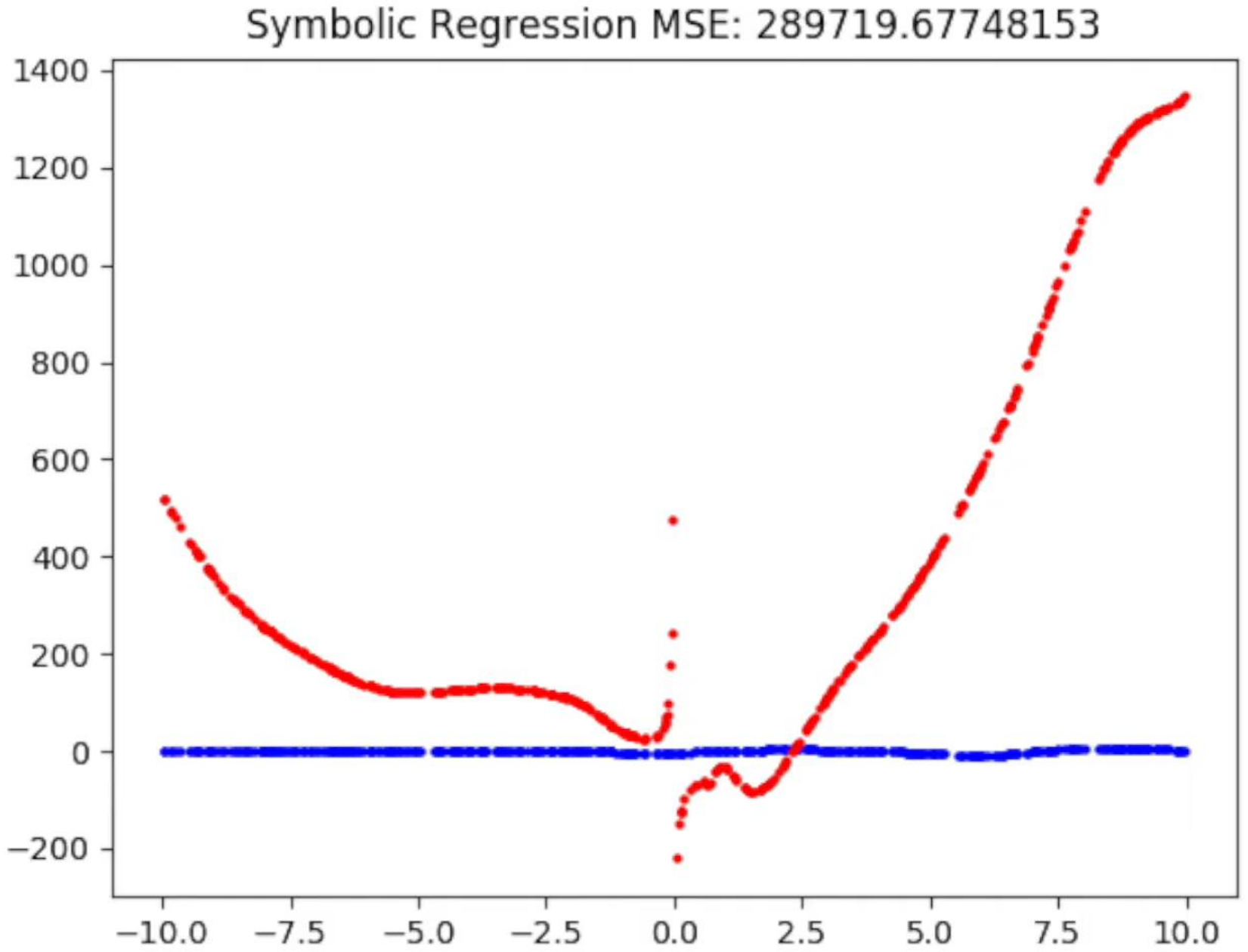
https://en.wikipedia.org/wiki/Early_stopping



https://en.wikipedia.org/wiki/Early_stopping

Diversity


40
20
0
Average edge diversity (%)
0
10
20
30
40
50
60
70
80
90
100
Number of generations
Without singular mating pool
With singular mating pool


Symbolic Regression MSE: 289719.67748153
1400
1200
1000
800
600
400
200
0
−200
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0

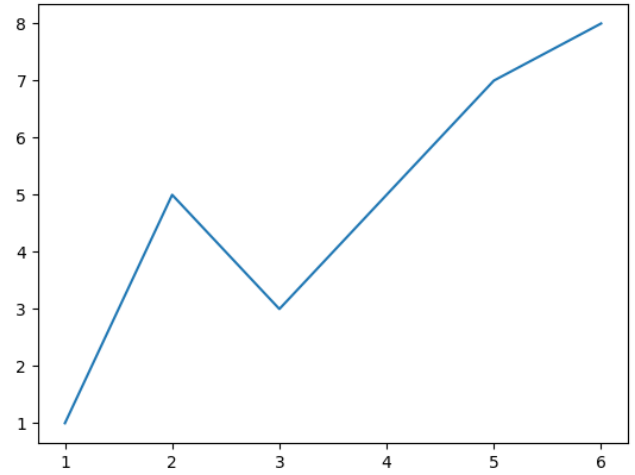
Max Segan

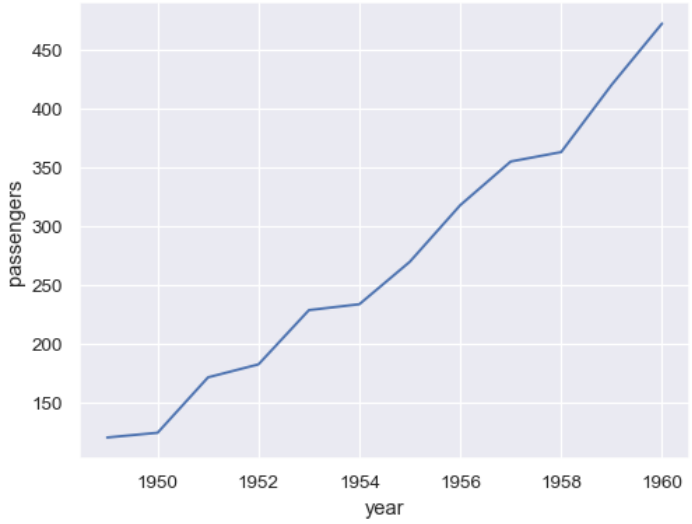
Tips

- Plot curves first. Don’t leave the curve-plotting to the end.
  - Learning curves can help you debug.
  - Always bring curves to office hours
- You can get max points even if you don’t solve the HW
  - Most of the grade is on process, not results
- Learn to use the cloud.
  - You can debug faster and work in parallel
- Develop the EA first (with crossover and mutation)
  - Disable the crossover and you have hill climber
  - Apply mutation to a blank solution and you have random search

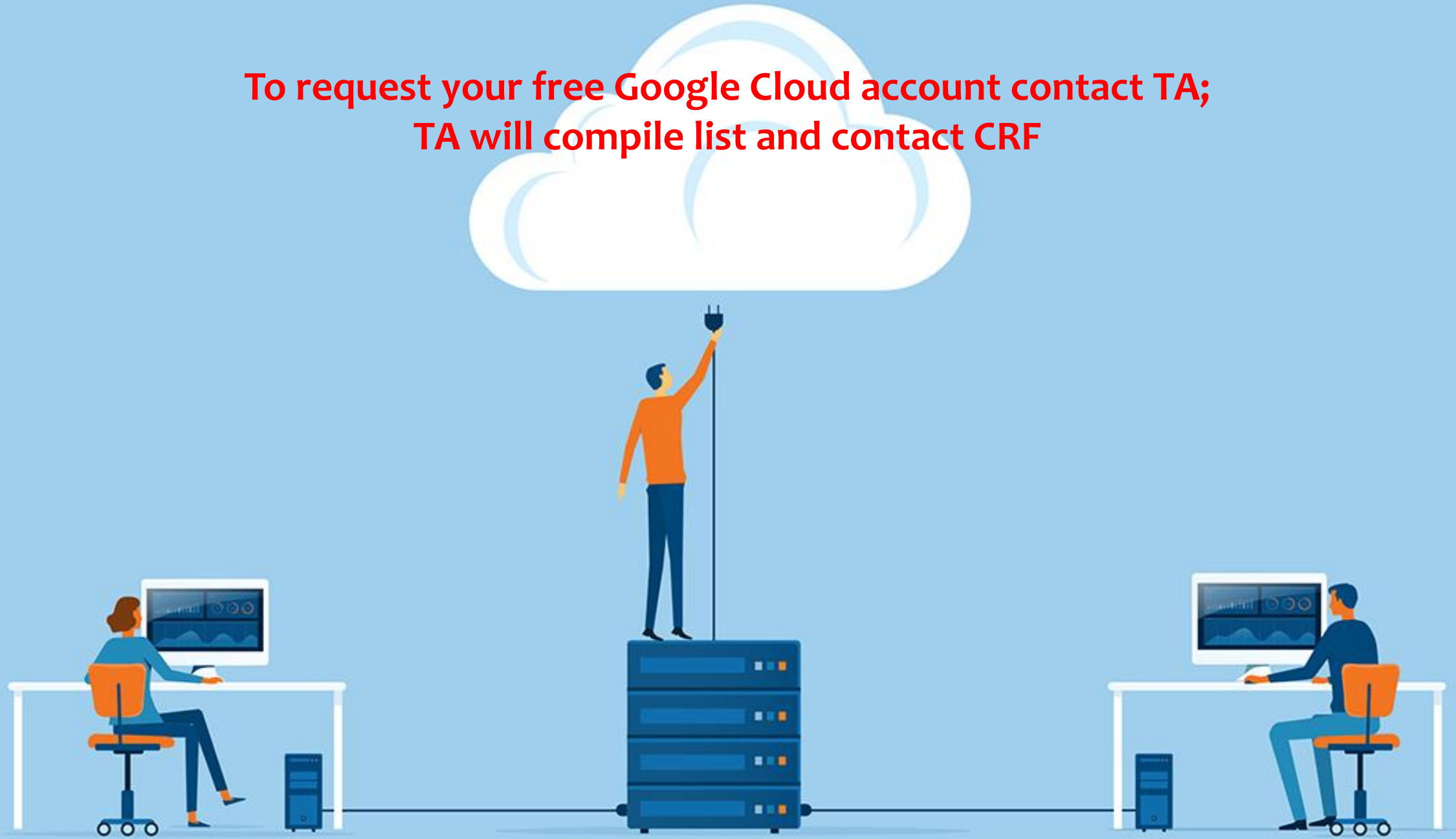
Charting in Python

- Matplotlib (https://matplotlib.org/)
  - Example: https://stackabuse.com/matplotlib-line-plot-tutorial-and-examples/
- Seaborn (https://seaborn.pydata.org/)
  - Example: https://seaborn.pydata.org/generated/seaborn.lineplot.html





To request your free Google Cloud account contact TA;
TA will compile list and contact CRF



https://youtu.be/cmLfMVorjgI