



ИНСТИТУТ ИНТЕЛЛЕКТУАЛЬНЫХ КИБЕРНЕТИЧЕСКИХ СИСТЕМ

Кафедра
«Криптология и кибербезопасность»

Отчет о домашней работе

«Биоинспирированные алгоритмы решения задач защиты информации.
Раздел Генетические алгоритмы. Д/З №1. Основы ГА.»

Исполнитель:
студент гр. Б17-505

Андрюшин М.А.

Преподаватель:

Борзунов. Г. И.

Москва – 2021

РЕФЕРАТ

Объектом исследования являются биоинспирированные алгоритмы решения задач защиты информации.

Цель работы — исследование генетических алгоритмов, задач, решаемых с помощью этих алгоритмов, оценка эффективности алгоритма.

В ходе работы рассмотрен, а также реализован базовый генетический алгоритм для решения задачи поиска максимума функции $\cos(\exp(x)) / \sin(\log(x))$ на отрезке $[2, 4]$, исследована зависимость времени поиска, числа поколений (генераций), точности нахождения решения от основных параметров генетического алгоритма, произведено сравнение найденного решения с действительным.

Результатом работы является описание и реализация базового генетического алгоритма, а также анализ зависимостей результатов работы алгоритма от основных параметров.

Определение

Генетический алгоритм (англ. genetic algorithm) — это эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём последовательного подбора, комбинирования и вариации искомых параметров с использованием механизмов, напоминающих биологическую эволюцию. Является разновидностью эволюционных вычислений (англ. evolutionary computation). Отличительной особенностью генетического алгоритма является акцент на использование оператора «скрещивания», который производит операцию рекомбинации решений-кандидатов, роль которой аналогична роли скрещивания в живой природе.

Кодирование пространства поиска

В данной работе применялось кодирование пространства поиска с помощью кода Грея, начальный отрезок $[2, 4]$ разбивался на 2^n частей, n может быть различным в зависимости от требуемой точности, каждый отрезок кодировался с помощью кода Грея.

Начальная популяция

Начальная популяция генерируется обычно случайно. Единственный критерий - достаточное разнообразие особей, чтобы популяция не свалилась в ближайший экстремум.

В рамках данной работы начальная популяция генерируется случайно.

Оператор отбора (селекции)

На этом этапе отбирается оптимальная популяция для дальнейшего размножения. Обычно берут определённое число лучших по

приспособленности. Имеет смысл также отбрасывать "клонов", т.е. особей с одинаковым набором генов.

Оператор скрещивания

При реализации использовался следующий алгоритм:

Задается число элитных особей, т.е. тех, которые не подвергаются скрещиванию, к остальным же применяется оператор скрещивания с заданной вероятностью. Оператор скрещивания одноточечный, т.е. берутся 2 особи, случайно определяется точка скрещивания и происходит обмен генами после этой точки.

Оператор мутаций

При реализации использовался следующий алгоритм:

К каждой особи с заданной вероятностью применяется оператор мутации, который заключается в том, что случайный бит хромосомы инвертируется.

Критерии останова

- выходом на «плато»;
- исчерпание числа поколений, отпущенных на эволюцию;
- исчерпание времени, отпущенного на эволюцию;
- исчерпание заданного числа обращений к целевой функции.

При реализации использовался следующий критерий:

Если значение функции не улучшается более чем X популяций, то алгоритм заканчивает работу.

Исследование зависимости результатов работы от основных параметров генетического алгоритма

Генетические алгоритмы богаты возможностями встраивания различных эвристик. До сих пор не существует (и не будет!) точных критериев оптимального размера популяции, способов мутаций и скрещивания, выбора начальной популяции и т.п. Поэтому данные параметры необходимо тщательно подбирать для каждой конкретной задачи. Далее будут приведены графики зависимости результатов работы от основных параметров генетического алгоритма.

Графики зависимости результатов работы алгоритма от вероятности кроссинговера, мутации

График зависимости времени поиска от вероятности кроссинговера, мутации

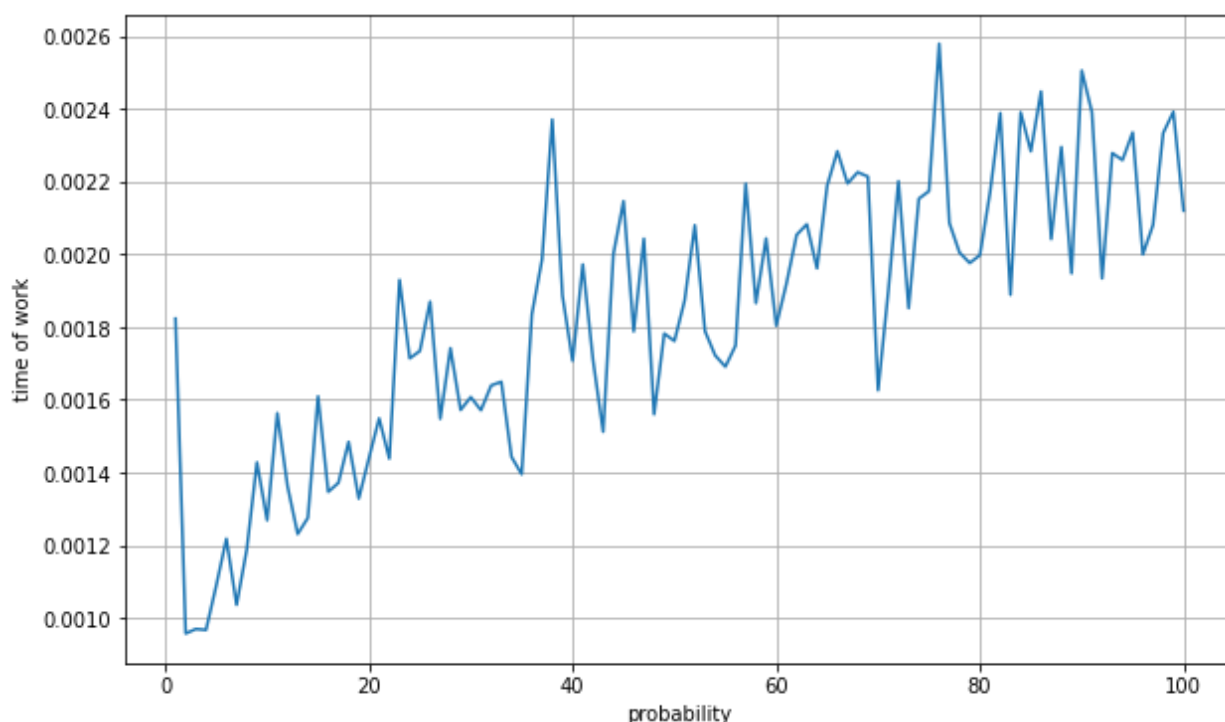


График зависимости числа поколений (генераций) от вероятности кроссинговера, мутации

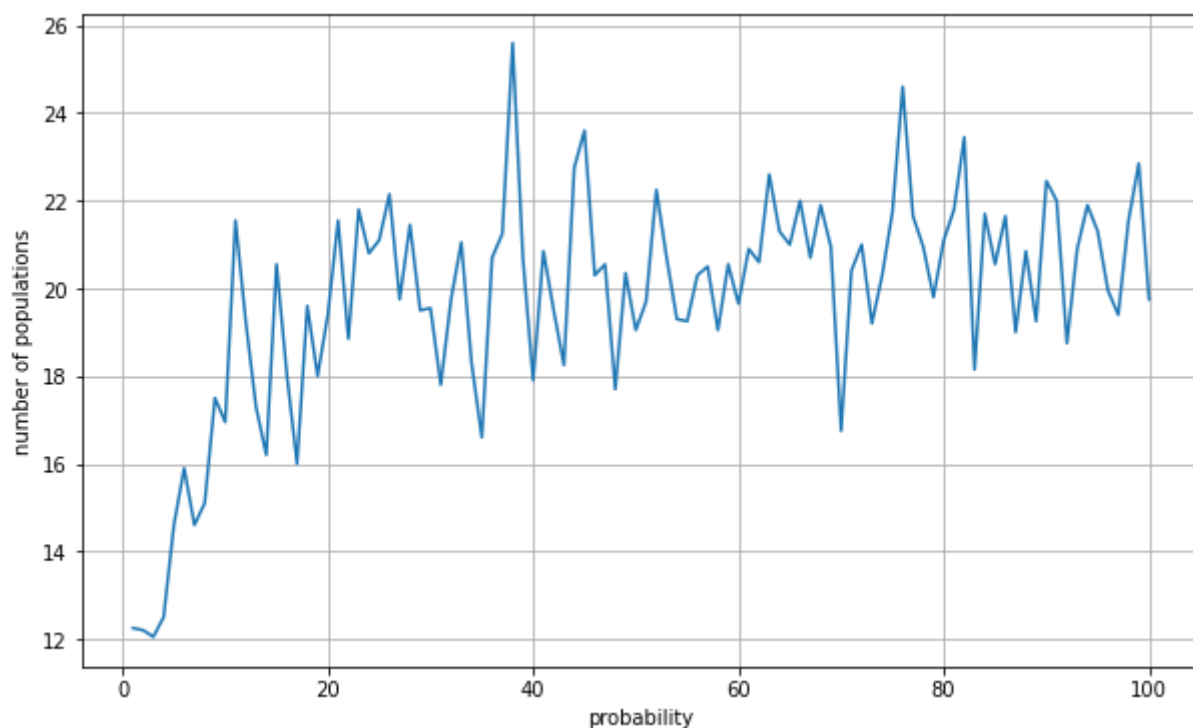
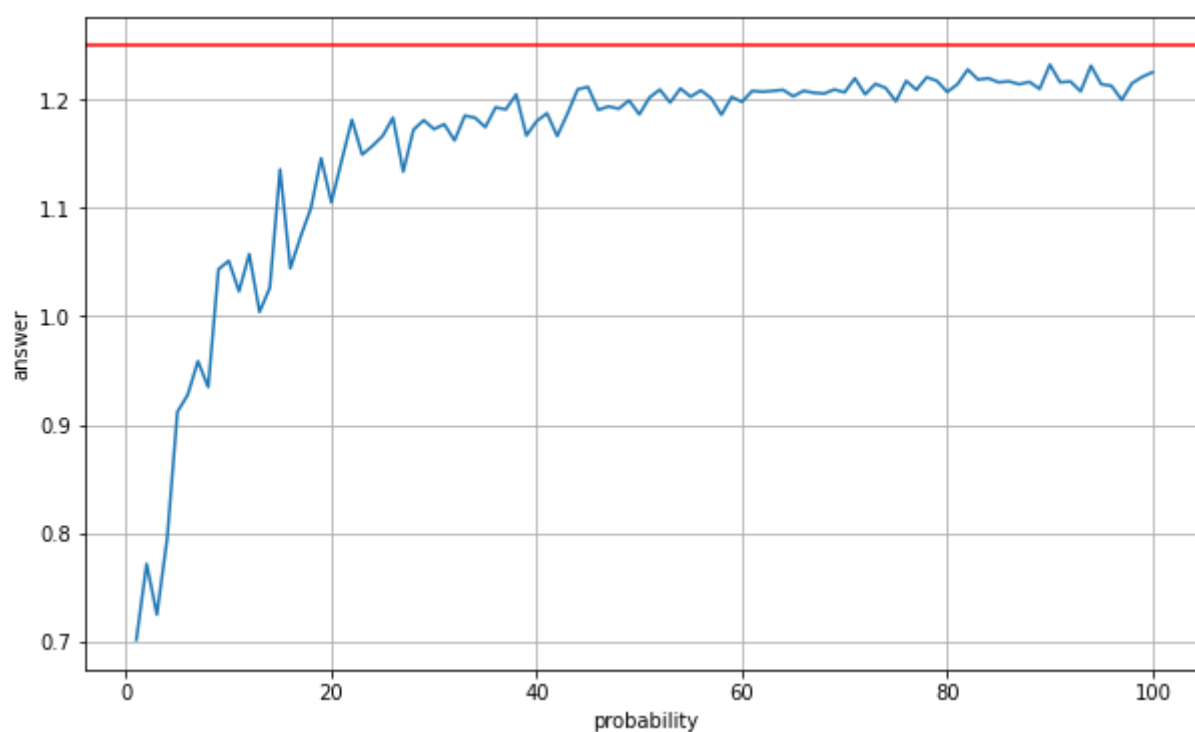


График зависимости точности решения от вероятности кроссинговера, мутации (красная линия обозначает максимальное значение функции)



Из этих графиков можно сделать следующие выводы:

- Время работы в среднем возрастает при увеличении вероятности кроссинговера и мутации, что в принципе можно объяснить тратой времени на применение самих этих операторов, а также тем, что с помощью них обнаруживаются “новые” области, где алгоритм может продолжить улучшение.
- Колебания на всех графиках говорят о том, что даже незначительное изменение параметра может сказаться на эффективности всего алгоритма, а также о том, что зависимость между параметрами достаточно сложна.
- Чем больше вероятности кроссинговера и мутации, тем более точный ответ получается.

Графики зависимости результатов работы алгоритма от количества особей в популяции

График зависимости времени поиска от количества особей в популяции

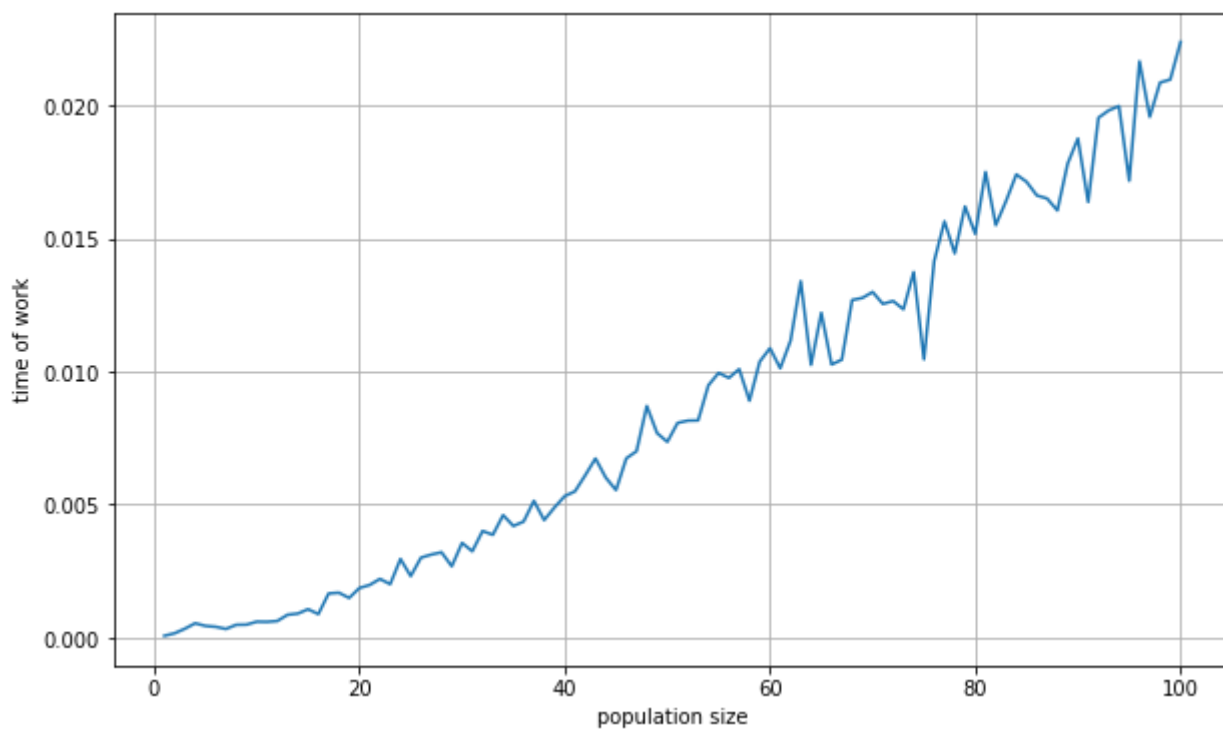


График зависимости числа поколений (генераций) от количества особей в популяции

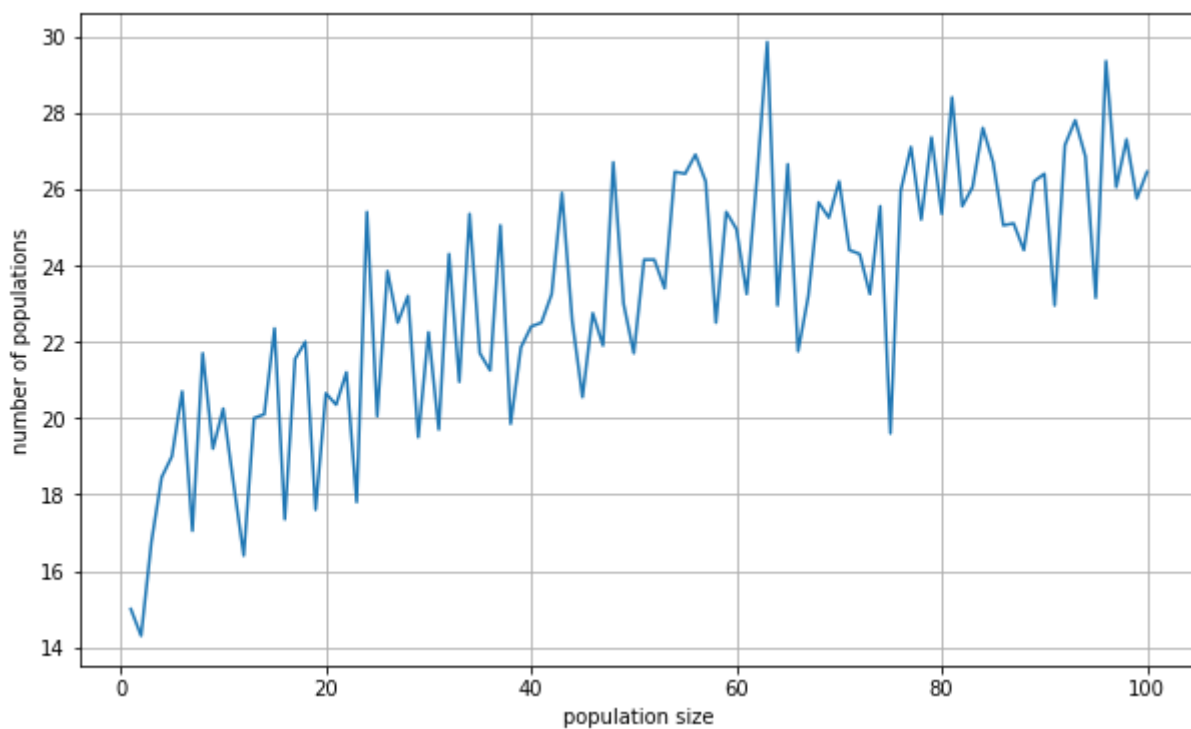
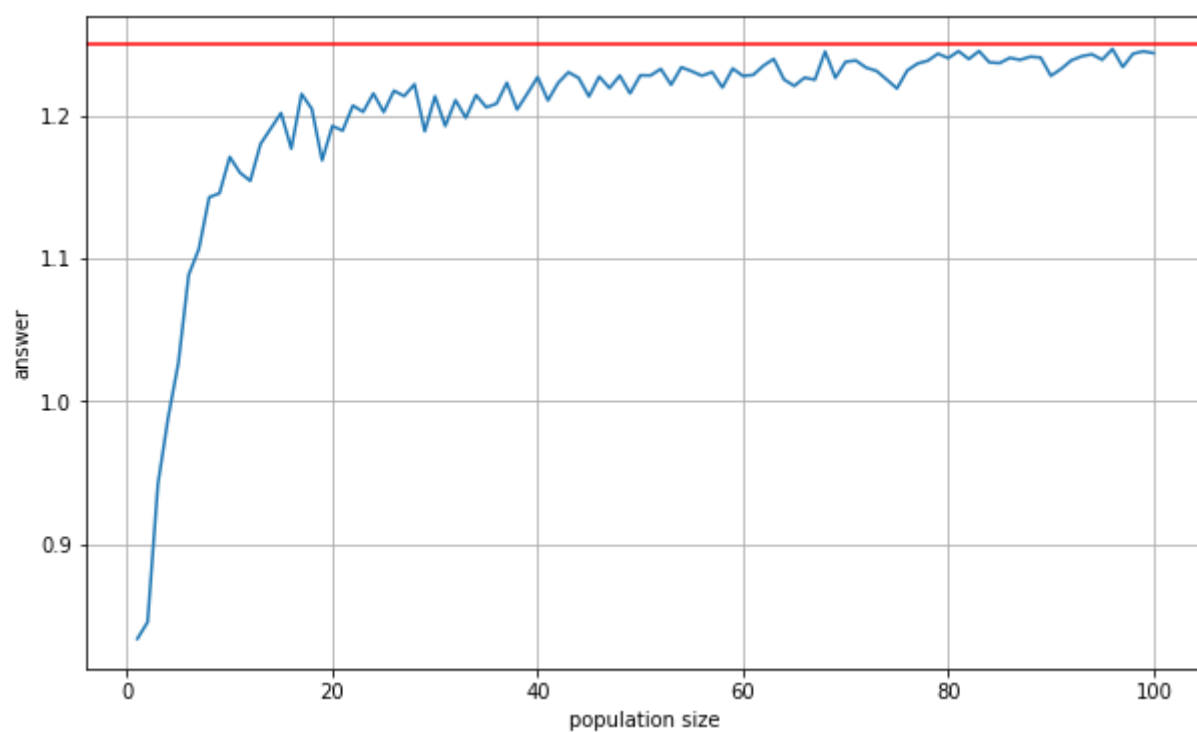


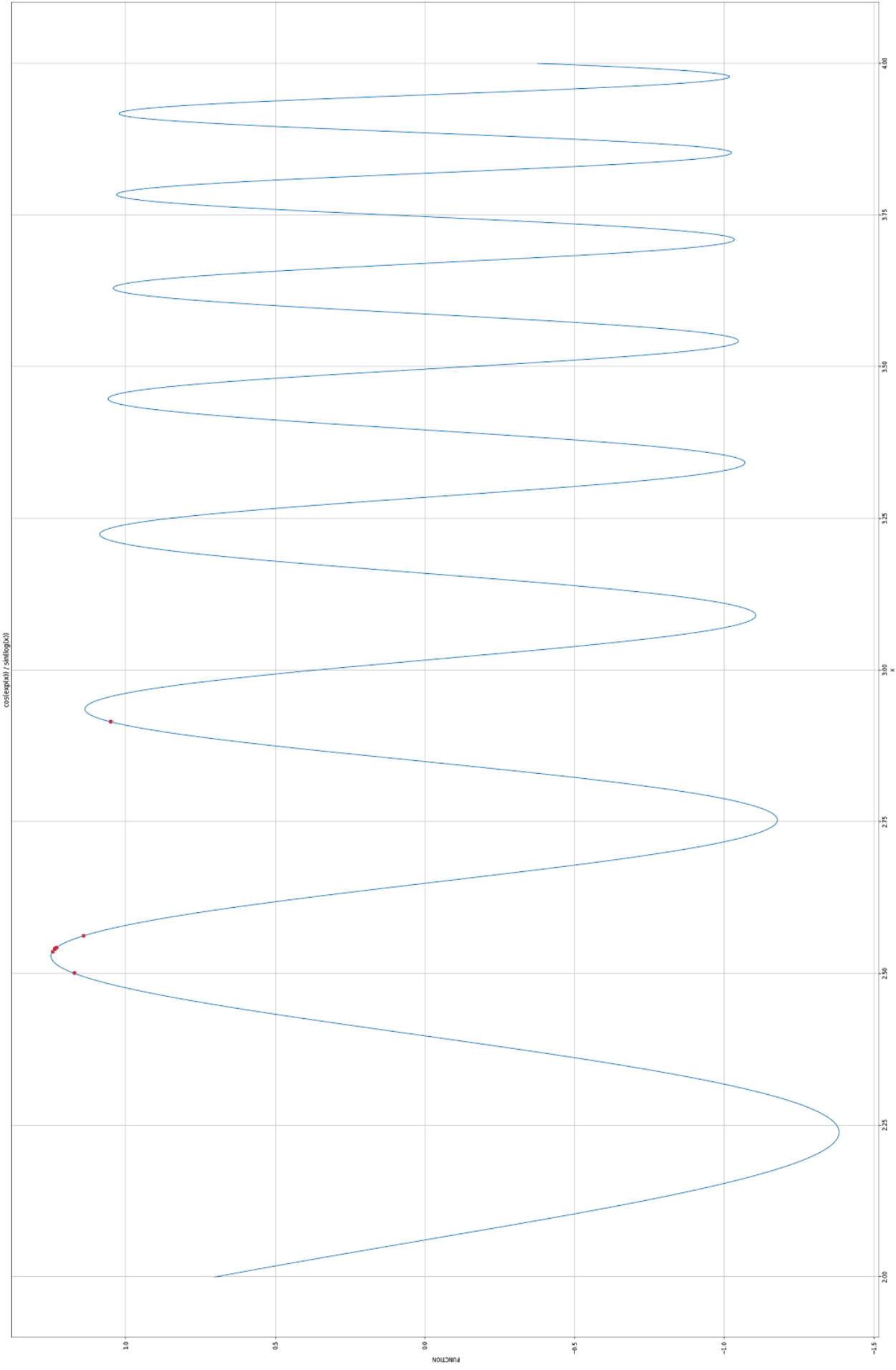
График зависимости точности решения от количества особей в популяции (красная линия обозначает максимальное значение функции)



Из этих графиков можно сделать следующие выводы:

- Время работы в среднем возрастает при увеличении вероятности кроссинговера и мутации, что в принципе можно объяснить тратой времени на поддержание размера популяции, так как на каждом этапе, необходимо применять операторы ко всем(почти всем) особям.
- Колебания на всех графиках говорят о том, что даже незначительное изменение параметра может сказаться на эффективности всего алгоритма, а также о том, что зависимость между параметрами достаточно сложна.
- Чем больше вероятности кроссинговера и мутации, тем более точный ответ получается.
- При достаточно большом количестве особей (>40) ответ становится достаточно близок к действительному.

График функции, на котором отмечены красным лучшими людьми каждой популяции



В результате работы найдены наиболее подходящие параметры генетического алгоритма, построены графики зависимостей, а также график, на котором отражены максимумы каждой популяции.

Максимальное полученное значение: 1.24630311.

Действительное максимальное значение при данной точности: 1.24910327.

Из этого можно сделать вывод, что глобальный максимум почти достигнут, ошибка составляет всего лишь тысячные доли.

Исходные коды и данные для построения графиков:

https://github.com/MaximAndryushin/genetic_alg/tree/master/hw1