Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

Ульяновский государственный Технический университет

Кафедра «Вычислительная техника»

Дисциплина «Системы искусственного интеллекта»

**Лабораторная работа №1**

**«Генетические алгоритмы»**

Выполнила:

студентка группы ИВТАСбд-41

Архипова Е.Ю

Проверил работу:

Хайруллин И. Д.

Ульяновск 2025

**Общее задание**

1. Необходимо разработать программу на языке python, реализующую генетический алгоритм по предложенному вариантом заданию.
2. Провести эксперименты по разным способам скрещивания (не менее 3-х), разным способам мутирования (не менее трех). Результат отобразить в виде графиков
3. Моделирование данных производить на основе максимально правдоподобных данных. Т.е. если рассматривается задача, в которой есть калорийность продуктов, то должны использоваться данные о реальных продуктах с реальной калорийностью.
4. Предоставить отчет о проделанной работе.

**Теоретические данные**

Генетические алгоритмы являются программной адаптацией эволюционного развития видов Чарльза Дарвина. Лучшая особь в популяции выбирается по принципу «выживает наиболее приспособленный».

Согласно принципам эволюции видов Чарльза Дарвина, адаптации генетических алгоритмов способны «развивать» решения реальных задач при соответствующем формальном описании. Существенным отличием генетических алгоритмов от реальной эволюции видов является моделирование лишь наиболее определяющих процессов.

**Основные принципы**

* Под популяцией понимается набор решений задачи.
* Каждое решение задачи оценивается по значению целевой функции или степени «приспособленности» для решения поставленной задачи.
* Особь в ГА характеризуется хромосомами, которые состоят из генов
* Как правило ген хранит 0 или 1 в зависимости от включения или не включении признака, характеризующего решение, в хромосому. Соответственно, хромосома – это битовая строка, описывающая решение.
* Но при решении реальных задач, применяются более сложные кодировки.

**Перед стартом ГА определяем:**

* Кодировку хромосомы (что будет представлять собой ген, и как гены будут участвовать в характеристике решения).
* Пространство гипотез (популяцию), из которых мы должны выбрать лучшую.
* Функцию приспособленности, оценивающую хромосомы.
* Набор и вид генетических операций (скрещивание, мутацию).
* Критерий остановки алгоритма (либо желаемое оптимальное значение, либо количество шагов эволюции популяции).

**Функция приспособленности**

Пример:

**=**

!!! Определение вида функции приспособленности – чрезвычайно важная задача, потому что от правильности ее определения будет во многом зависеть успешность решения.

**ГА: Мутации**

Если говорить формально, то мутация – это изменение одного или нескольких генов хромосомы вследствие случайного влияния.

Мутация – это изменение значения одного или нескольких генов хромосомы на противоположный или четко заданный.

**ГА: Скрещивание**

Операция скрещивания, или кроссовер, – операция, которая получает из двух хромосом одну, используя заданную маску. По сути из каждой хромосомы «вырезается» кусок, который помещается в новую. Существует несколько видов кроссовера.

*Одноточечный кроссовер:* «разрез» хромосомы происходит только в одной точке, и новая особь получается путем соединения первой части первой хромосомы и второй части второй хромосомы.

*Двухточечный кроссовер:* есть две точки «разреза», и новая хромосома получается из двух частей первой хромосомы и одной части второй.

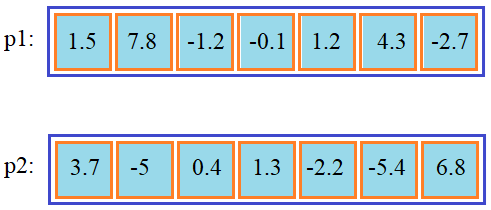
**Равномерное скрещивание**

Формирование потомков из скрещивания отдельных пар ген родителей, которые отбираются из хромосом случайным образом. Здесь слово «равномерное» означает использование датчика равномерного распределения для определения случайной позиции генов в хромосоме.

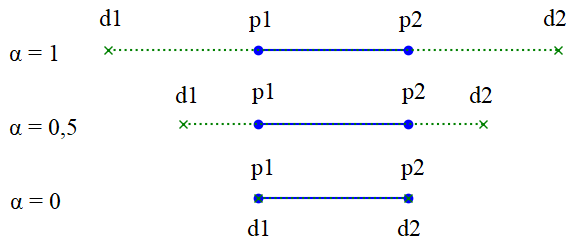


**Скрещивание смешением**

blend crossover – BLX



перебираем соответствующие пары ген этих двух родителей и вычисляем интервал для последующей генерации случайных значений генов потомков.



Далее, в интервале [d1; d2] случайным образом выбираются два числа: для первого и второго потомков. И так для всех генов. На выходе получаем двух потомков с новым набором ген.



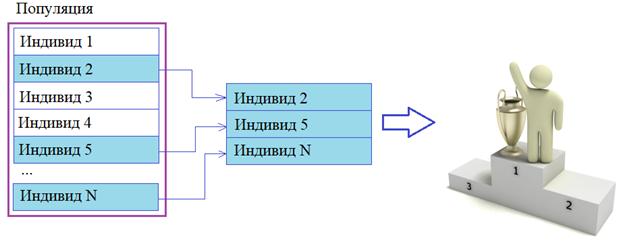
**Генерация нового поколения**

Обычно в популяцию выбирается s наиболее приспособленных особей, которые и дают в ней потомство. Для того чтобы выбрать хромосомы в популяцию, можно использовать разные методы:

* турнирный,
* ранговый
* метод рулетки
* стохастическая универсальная выборка
* масштабирование приспособленности

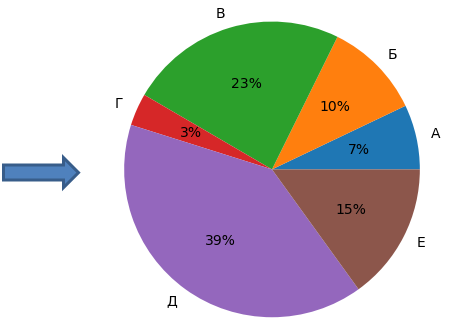
**Турнирный отбор**

Каждый раз из популяции случайным образом отбирается несколько претендентов (от двух и более). Затем, среди отобранных участников выбирается наиболее приспособленный (с наибольшим значением функции принадлежности). Он и переходит в новую выборку.



Процесс повторяется до тех пор, пока число «родителей» не станет равно размеру популяции.

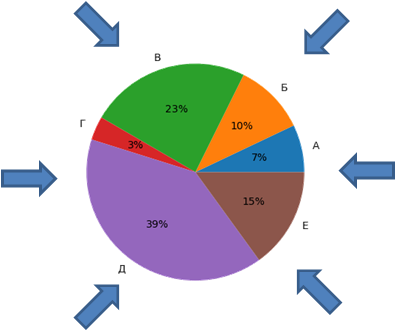
**Правило рулетки**

****

*отбор пропорциональной приспособленности* (fitness proportionate selection– FPS) раскручиваем круг и тот сектор, на который будет указывать стрелка, будет выбран. Таким образом, отбирается особь в качестве родителя. Очевидно, что чаще (с наибольшей вероятностью) будут отбираться индивидуумы с большей приспособленностью, так как у них сектор занимает большую долю.

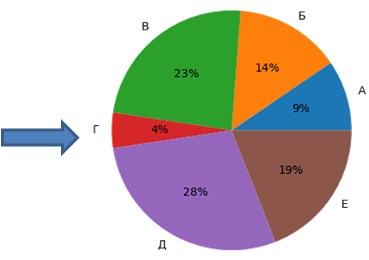
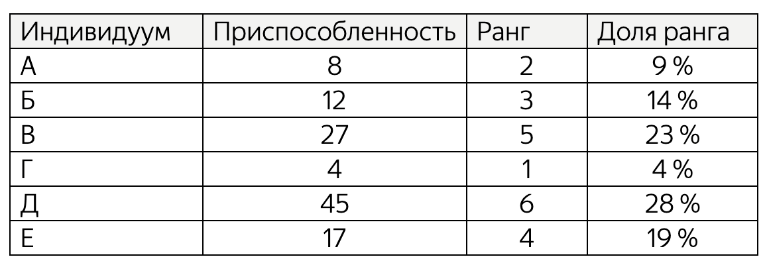
**Стохастическая универсальная выборка**

Преимущество этого подхода в том, что он гарантирует отбор не только индивида с большой долей сектора, но, скорее всего, будут отобраны и другие особи с меньшими секторами. То есть, выбор более равномерен и сохраняет, в некоторой степени, разнообразие популяции.



(stochastic universal sampling – SUS) – модифицированный вариант рулетки, когда точки отбора (стрелки) располагаются равномерно вокруг диаграммы

**Ранжированный отбор**

****

Вначале все индивидуумы ранжируются (упорядочиваются) по возрастанию их приспособленности.

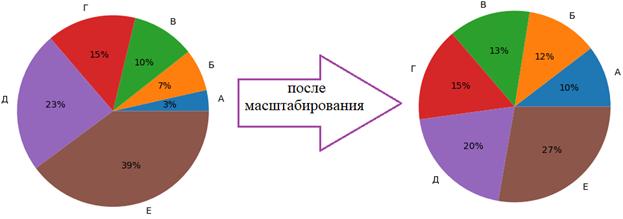
Затем, вычисляются доли относительно ранга (а не приспособленности, как в правиле рулетки) и, тем самым, доли секторов становятся более равномерными, а значит, более равномерно будут отбираться родители из популяции.

**Масштабирование приспособленности**

****

И мы все эти значения от 4 до 45 хотим перевести в новый диапазон [30; 80]. Это делается простой математической операцией:

** **

****

**Общая схема ГА**

****

**Описание набора данных**

Для реализации генетического алгоритма подбора оптимального рациона питания были использованы данные из двух источников.

База данных о пищевой ценности продуктов была взята с Kaggle (Food Nutrition Dataset) (<https://www.kaggle.com/datasets/shrutisaxena/food-nutrition-dataset> ). Из исходного набора были выбраны ключевые колонки: название продукта, калорийность, содержание белков, жиров и углеводов. Эти данные были предобработаны - значения питательных веществ умножены на 10 для приведения к реальным величинам и округлены до одного знака после запятой.

Данные о ценах на продукты были получены с сайта Федеральной службы государственной статистики РФ([https://www.fedstat.ru/indicator/31448?utm\_source=chatgpt.com#](https://www.fedstat.ru/indicator/31448?utm_source=chatgpt.com) ). Была выбрана Ульяновская область, последние доступные данные за март 2025 года. Из таблицы были взяты розничные цены на продовольственные товары.

Для объединения информации был создан словарь соответствия между английскими названиями продуктов из базы Kaggle и русскими названиями из данных Росстата. После этого выполнено объединение таблиц по названиям продуктов, что позволило создать единый датасет, содержащий как пищевую ценность, так и цены.

Итоговый набор данных содержит 1400 продуктов с полной информацией о калорийности, содержании белков, жиров, углеводов и актуальными ценами.

Для определения медицинских норм питания использовались Методические рекомендации MP 2.3.1.0253-21 "Нормы физиологических потребностей в энергии и пищевых веществах для различных групп населения Российской Федерации", утвержденные Федеральной службой по надзору в сфере защиты прав потребителей и благополучия человека 22 июля 2021 года.( <https://www.garant.ru/products/ipo/prime/doc/402716140/#5033> )

Поскольку в рекомендациях нормы разделены по полу, было принято решение взять нормы для мужчин по нижней границе.

Так как продукты в наборе данных представлены в крупных объемах (килограммы, литры), расчет выполняется для недельного рациона.

Были установлены следующие суточные нормы:

* Калорийность: 2150 ккал/сутки \* 7
* Белки: 75 г/сутки \* 7
* Жиры: 72 г/сутки \* 7
* Углеводы: 301 г/сутки \* 7

**Описание реализации**

В реализации генетического алгоритма использовалась бинарная кодировка хромосомы. Хромосома представляет собой вектор длиной N (где количество продуктов бд), в котором каждый ген соответствует одному продукту. Значение гена 1 означает, что продукт включен в рацион, 0 - не включен. Таким образом, хромосома полностью описывает состав продуктовой корзины.

Популяция состоит из 100 хромосом, каждая из которых инициализируется случайным образом с фиксированным количеством выбранных продуктов K = 10. Это означает, что в каждом начальном рационе содержится ровно 10 различных продуктов из базы данных.

Функция приспособленности оценивает качество хромосомы по двум основным критериям: стоимости рациона и соответствию медицинским нормам питания. Функция вычисляется как сумма общей стоимости выбранных продуктов и штрафных санкций за отклонение от норм по калорийности, белкам, жирам и углеводам. Штрафы рассчитываются как относительные отклонения от норм, умноженные на коэффициент 10000, что обеспечивает приоритетность соблюдения медицинских требований над минимизацией стоимости.

Для отбора особей используется турнирный отбор размером 3. На каждом шаге случайным образом выбираются 3 особи из популяции, и лучшая из них (с наименьшим значением функции приспособленности) проходит в следующее поколение. Этот процесс повторяется до формирования новой популяции того же размера.

Были реализованы три метода скрещивания:

* одноточечное
* двухточечное
* равномерное скрещивание

После операции скрещивания выполняется восстановление хромосомы для сохранения фиксированного количества продуктов. Также реализованы три метода мутации:

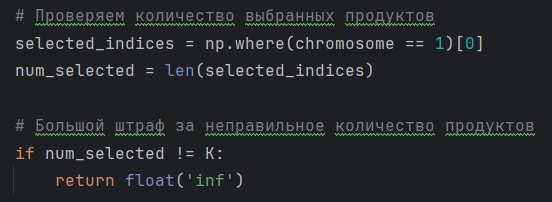
*Точечная мутация* - изменение значения одного случайного гена на противоположное. Данный метод обеспечивает локальный поиск вблизи текущего решения и вносит минимальные изменения в состав рациона.

*Множественная мутация* - одновременное изменение значений нескольких случайных генов (от 1 до 3). Этот метод позволяет исследовать более широкую область пространства поиска и способствует выходу из локальных оптимумов.

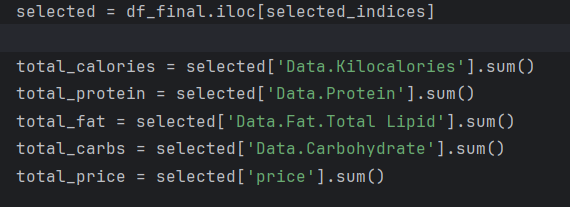
*Улучшающая мутация* - целенаправленная замена самого дорогого продукта в рационе на более дешевый аналог из той же категории. Данный метод использует информацию о качестве решения для осознанного улучшения стоимости рациона.

Критерием остановки алгоритма является достижение 200 поколений эволюции. Такой подход обеспечивает предсказуемое время выполнения и позволяет алгоритму достаточно времени для сходимости к удовлетворительному решению. После завершения 200 поколений алгоритм возвращает лучшее найденное решение.

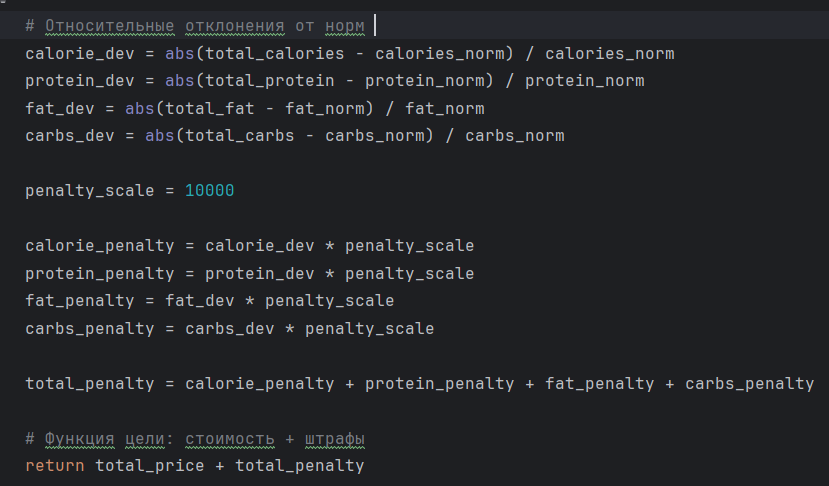
Функция приспособленности реализована следующим образом:



Функция начинается с проверки корректности хромосомы - количества выбранных продуктов должно быть ровно K. Нарушение этого условия приводит к максимальному штрафу.

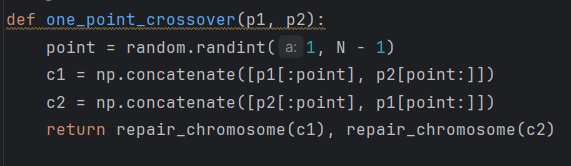


Для выбранных продуктов вычисляются суммарные значения питательных веществ и общей стоимости.



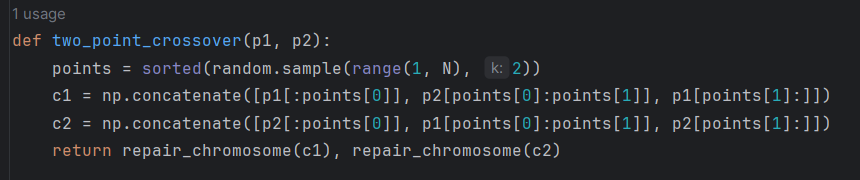
Рассчитываются относительные отклонения от медицинских норм, которые умножаются на коэффициент 10000 для обеспечения приоритета соблюдения норм над минимизацией стоимости.

Одноточечное скрещивание



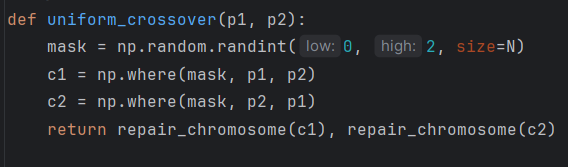
Хромосомы родителей разделяются в случайной точке, после чего хвостовые части обмениваются. Результат восстанавливается до K продуктов.

Двухточечное скрещивание



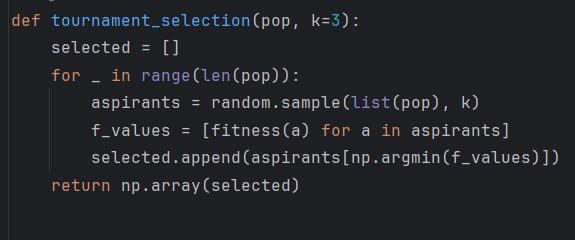
Выбираются две случайные точки, средние сегменты хромосом обмениваются между родителями.

Равномерное скрещивание



Для каждого гена случайно выбирается родитель, от которого он наследуется.

Турнирный отбор



Для каждой особи следующего поколения случайно выбираются k=3 кандидата, из которых отбирается особь с наилучшим значением функции приспособленности.

Алгоритм выполняет 200 поколений, на каждом этапе применяя отбор, скрещивание и мутацию для создания новой популяции.

**Вывод**

В ходе выполнения лабораторной работы был успешно разработан и реализован генетический алгоритм для решения задачи оптимизации рациона питания. Основной целью работы было нахождение самого дешевого набора из продуктов, удовлетворяющего заданным медицинским нормам по калорийности, содержанию белков, жиров и углеводов.