

КОГНИТИВНЕ МОДЕЛЮВАННЯ

122 «Комп'ютерні науки»

КНМ-21

2021 / 2022 навчальний рік

ЕВОЛЮЦІНЕ МОДЕЛЮВАННЯ. ЕВОЛЮЦІЙНІ ТА ГЕНЕТИЧНІ АЛГОРИТМИ.

Генетичні алгоритми

1. *Створення популяції*
2. *Селекція*
3. *Схрещення*
4. *Мутація*
5. *Редукція*

Задача оптимізації

Стандартная постановка:

Задано:

- Допустимое множество независимых
- переменных $X = \{\vec{x} | g_i(\vec{x}) \leq 0, i = 0, 1, \dots, m\} \in \mathbb{R}^n$
- Целевая функция – отображение $f: X \rightarrow \mathbb{R}$
- Критерий поиска (***min*** или ***max*** целевой функции)

Решение: найти такое $\vec{x}^* \in X$, что

$$f(\vec{x}^*) = \min_{\vec{x} \in X} f(\vec{x})$$

Решением задачи занимается теория
математического программирования.

Общая идея генетических алгоритмов



1. Генерируется популяция особей (набор допустимых x)
2. Селекция (reproduction) - отбираются претенденты - для них вычисляется f (*fitness function*). Обираются некоторой вероятностью лучшие (!родители).
3. Отобранные особи (! родители) скрещиваются (*crossover*) – новые особи x наследуют свойства родителей.
4. Мутация (*mutation*) – случайные изменения признаков - x
5. Формируется новая популяция

Процесс повторяется – множество **поколений**, пока не будет достигнут приемлемый результат

Создание исходной популяции

Хромосома, представляющая из себя набор генов - n разрядов.

«Одеяло» – формирование полной популяции , содержащей все возможные хромосомы $\rightarrow 2^n$ вариантов.

«Дробовик» – формирование достаточно большого **случайного** подмножества вариантов.

«Фокусировка» – формирование подмножества решений, включающих разновидности одного решения.

Чаще всего - второй способ \rightarrow в результате эволюции есть возможность перейти в другие подобласти поиска и каждая из них имеет сравнительно небольшое пространство поиска.

Селекция - отбор родителей

В результате отбора из текущей популяции порождается промежуточная популяция – потенциальных родителей.

«Рулетка» – пропорциональный отбор особи (решения) отображаются в отрезки линии (или сектора рулетки) таким образом, что их размер пропорционален значению целевой функции $f()$ для данной особи.

Далее нужное количество раз случайно генерируются числа из диапазона $[0,1]$ и в промежуточную популяцию выбираются те особи, в чей "отрезок" попадают эти случайные числа. Таким образом, каждая попытка представляет собой случайное число из отрезка $[0,1]$, в результате чего выбирается особь, соответствующая выбранному отрезку.

Селекция - отбор родителей

«Ранжирование» –

особи (решения) сортируются согласно значениям целевой функции $f()$. Пусть N особей.

Определяется номер i в списке. Необходимое количество раз случайно генерируются целое число из диапазона $[0, N-1]$ и в промежуточную популяцию выбираются особи этим индексом.

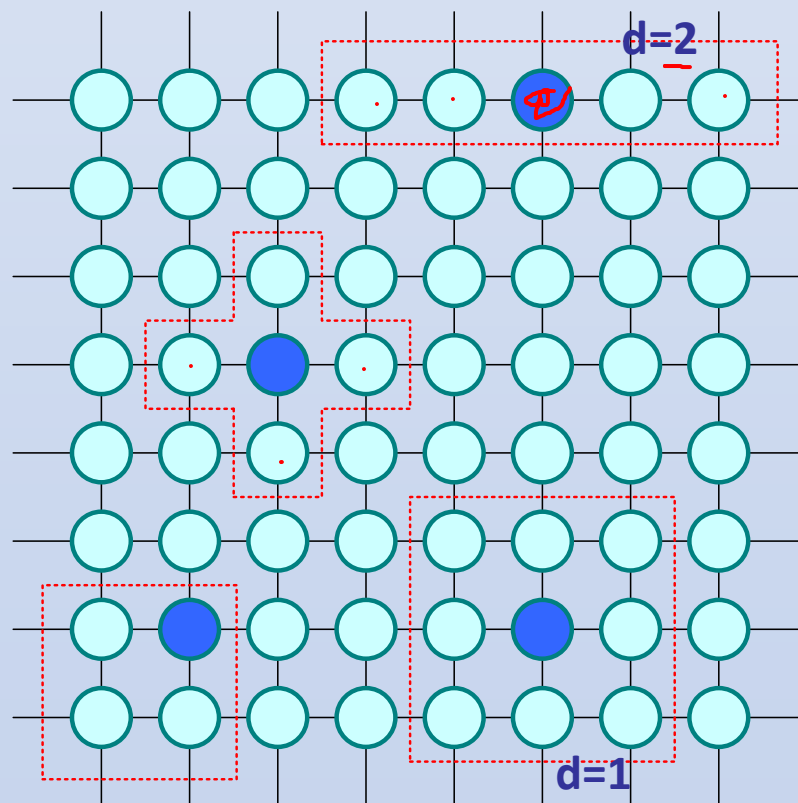
Вероятность отбора особи зависит только от ее позиции (номера) в этом упорядоченном множестве особей, а не от самого значения целевой функции.

Селекция - отбор родителей

«Локальный отбор» –

отбор производится среди особей, которые находятся в некоторой ограниченной среде, где определено отношение соседства. При этом в качестве соседей подразумевается множество партнеров для операции скрещивания.

- Одномерное, линейное соседство
- Двумерное, четырехсвязное соседство
- Двумерное, восьмисвязное соседство



Селекция - отбор родителей

«Отбор усечением» –

особи популяции упорядочиваются согласно их значениям целевой функции (ранжирование). Затем в качестве родителей выбираются только **лучшие** особи. Далее, с равной вероятностью, среди них случайным образом выбирают пары, которые производят потомков.

При этом методе используется параметр T – порог отсечения (интенсивность отбора) – определяет часть популяции для отбора в качестве родителей $10\% < T < 50\%$.

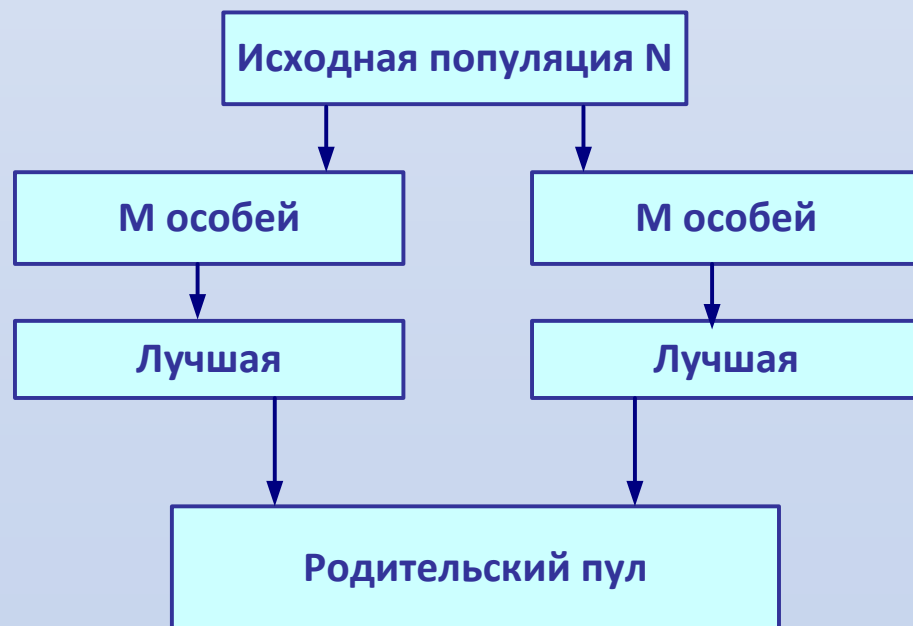
Селекция - отбор родителей

«Турнир» –

все особи популяции разбиваются на подгруппы размера M , с последующим выбором в каждой из них особи с лучшим значением фитнес-функции.

M – размер тура $1 < M < N$,
чаще всего $M = 2, M = 3$.

Экспериментальные исследования показывают, что часто турнирный метод *эффективней* метода рулетки.



Выбор пар для скрещивания

Случайный выбор (панмиксия)

оба родителя случайным образом выбираются из всей промежуточной популяции.

При этом любая особь может входить в несколько пар.

Метод является универсальным для решения различных задач, но достаточно критичен к численности популяции, поскольку его эффективность снижается с ростом размера популяции N .

Селективный выбор

Родителями могут стать только те особи, значение целевой функции которых не меньше среднего значения по популяции при равной вероятности этих кандидатов составить брачную пару.

Такой подход обеспечивает более быструю сходимость алгоритма, но неприемлем для мультимодальных задач (имеющих несколько экстремумов), для которых этот метод, как правило, быстро сходится к одному из решений.

Выбор пар для скрещивания

Инбридинг

первый член пары выбирается случайно, а вторым является максимально близкая к нему особь $d < d^G$.

Аутбридинг

пары формируются из максимально далеких особей $d > d^H$.

Степень родства – **расстояние** между хромосомами – расстояние между **кодами**.
Простейшее определение – расстояние Хэмминга – число различных генов хромосомах особей.

Эффективным является использование на начальных этапах работы алгоритма инбридинга, а на заключительных этапах – аутбридинга.

1	0	1	0	0	1	1	1
1	1	0	0	0	1	1	0

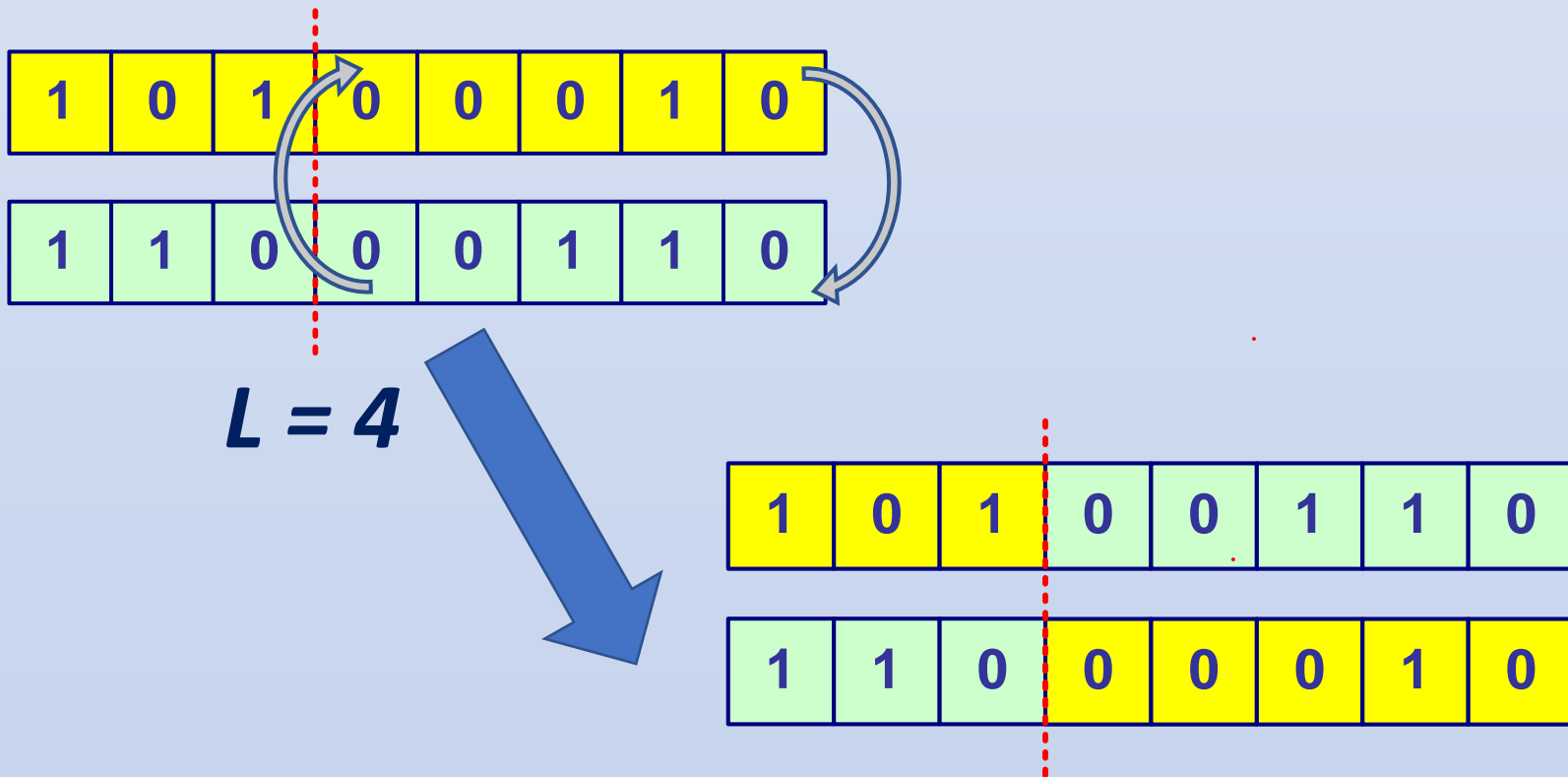
$d=3$

Скрещивание (кроссинговер)

Бинарное скрещивание

Одноточечный кроссинговер

Случайным образом выбирается точка скрещивания и производится обмен фрагментами после (перед) точки скрещивания L .

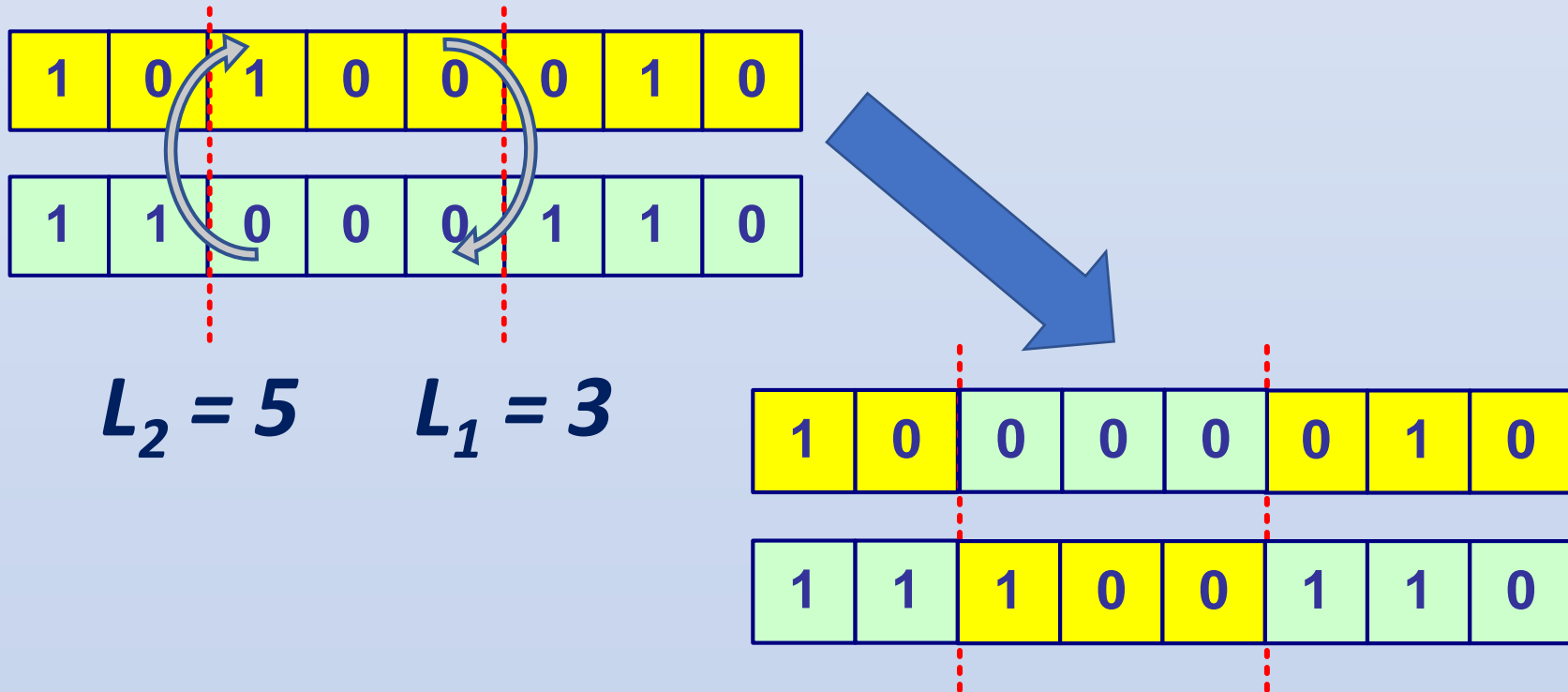


Скрещивание (кроссинговер)

Бинарное скрещивание

Многоточечный кроссинговер

Случайным образом выбирается несколько точек L_k , устанавливается правило обмена фрагментами производится обмен.



Скращивание (кроссинговер)

Бинарное скращивание

Однородный кроссинговер

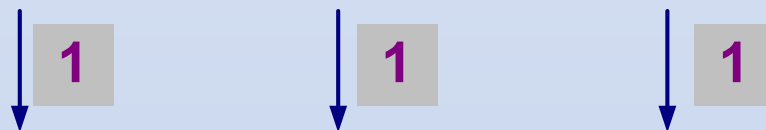
Каждый ген потомка создается путем копирования соответствующего гена из первого или второго родителя → есть каждая позиция потенциально является точкой кроссинговера.

1	0	0	1	0	0	1	0
---	---	---	---	---	---	---	---

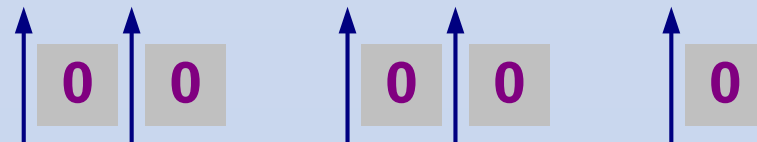
Случайным образом генерируется двоичная маска кроссинговера той же длины что у хромосом родителей.

Четность бита маски показывает родителя, из которого копируется ген потомка.

1	0	1	0	0	0	0	1
---	---	---	---	---	---	---	---



1	1	0	0	0	1	0	1
---	---	---	---	---	---	---	---



0	1	0	0	0	1	1	1
---	---	---	---	---	---	---	---

Мутация

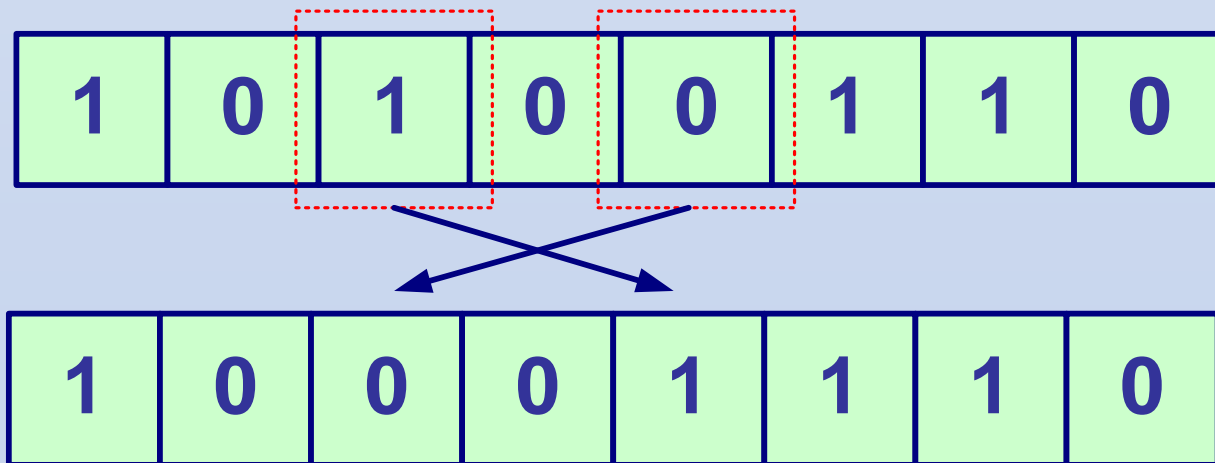
Бинарная мутация

Классическая мутация.

Для каждой особи случайным образом выбирается позиция и с малой вероятностью ($0,001 < p < 0,01$) выполняется инвертирование в этой позиции.

Инверсная мутация.

случайным образом выбираются две позиции в особи и с малой вероятностью выполняется обмен генами в этих позициях.



Редукция

Устранение плохих особей (решений), полученных в процессе формирования новой популяции.

Поколение t , особей $r(t) = N$.
Скращивание добавило $r_s(t)$ особей.
Мутация добавила $r_m(t)$ особей .

Численность репродукционной группы поколения t

$$R(t) = r(t) + r_s(t) + r_m(t)$$

В новом поколении число особей д.б. = N
Надо устранить плохих.

Редукция

Чистая замена. С помощью скрещивания и мутации генерируется ровно столько особей, сколько было родителей. Родитель живет одно поколение и отбрасывается.

Элитарная схема. Потомков генерируется меньше, чем было родителей. Вновь построенные потомки заменяют худших родителей согласно значениям фитнес-функции.

Равномерная случайная замена. Потомков генерируется меньше, чем было родителей. Случайным способом устраняется столько родителей, сколько сгенерировано потомков.

Редукция

Пропорциональная редукция. Потомков генерируется больше чем родителей. Заданное число родителей заменяется лучшими потомками .

Селекционная схема. Потомки и родители равны. Все ранжируются по значению фитнес функции и отбирается N лучших.

Равномерная случайная замена. Потомков генерируется меньше, чем было родителей. Случайным способом устраняется столько родителей, сколько сгенерировано потомков.

ГА + ИНС

1. Генетические алгоритмы при обучении НС

ИНС: Задача настройки – поиск таких синаптических весов $w_{ij}^{(l)}$, что функция ошибки

$$E(w_{ij}) = 0.5 \sum e_j^2(w_{ij}),$$

имеет минимальное значение.

ГАС: задача оптимизации

- Переменные \mathbf{x} есть веса w_{ij}
- Целевая функция $f()$ есть $E(k)$
- Критерий поиска - ***min*** целевой функции

Эксперименты показывают, что в некоторых случаях ГА лучший метод обучения НС, чем обратное распространение.

ГА + ИНС

2. Генетические алгоритмы оптимизации архитектуры НС.

Создать популяцию из нескольких ИНС.

Присвоить всем ИНС случайные (в некотором диапазоне) гиперпараметры (количество слоев, количество нейронов, связи).

Цикл:

- Обучить все NN одновременно или по одному.

- Рассчитать стоимость обучения.

- Рассчитать "фитнес функцию" (насколько ИНС справилась с этой итерацией), исходя из ее стоимости. Фитнес используется для повышения шансов "воспроизводства" ИНН.

ГА + ИНС

2. Генетические алгоритмы оптимизации архитектуры ИС.

Найти максимальное значение фитнеса среди поколения.

Выбрать 2 ИНС на основе системы вероятностей относительно их пригодности.

Скрестите гены 2 ИНС.

Мутировать гены дочерней ИНС.

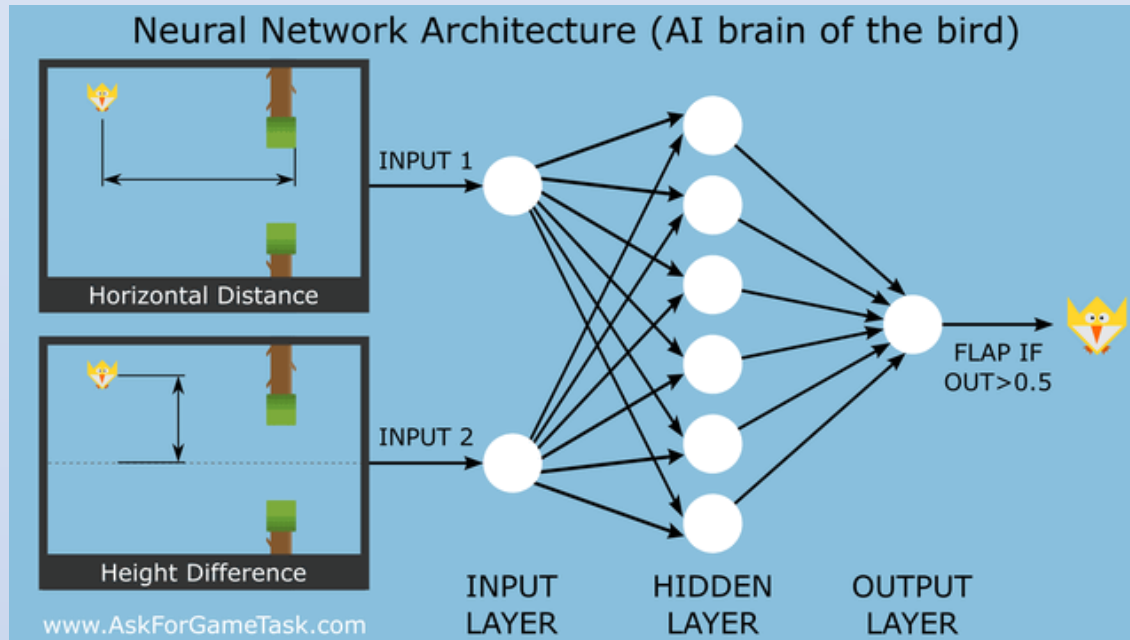
Сформировать новое поколение ИНС

Цикл повторяем пока не найдем удовлетворительной архитектуры ИНС.

ГА + ИНС

Пример. Игра FlappyBird

<http://flappybird.io/>

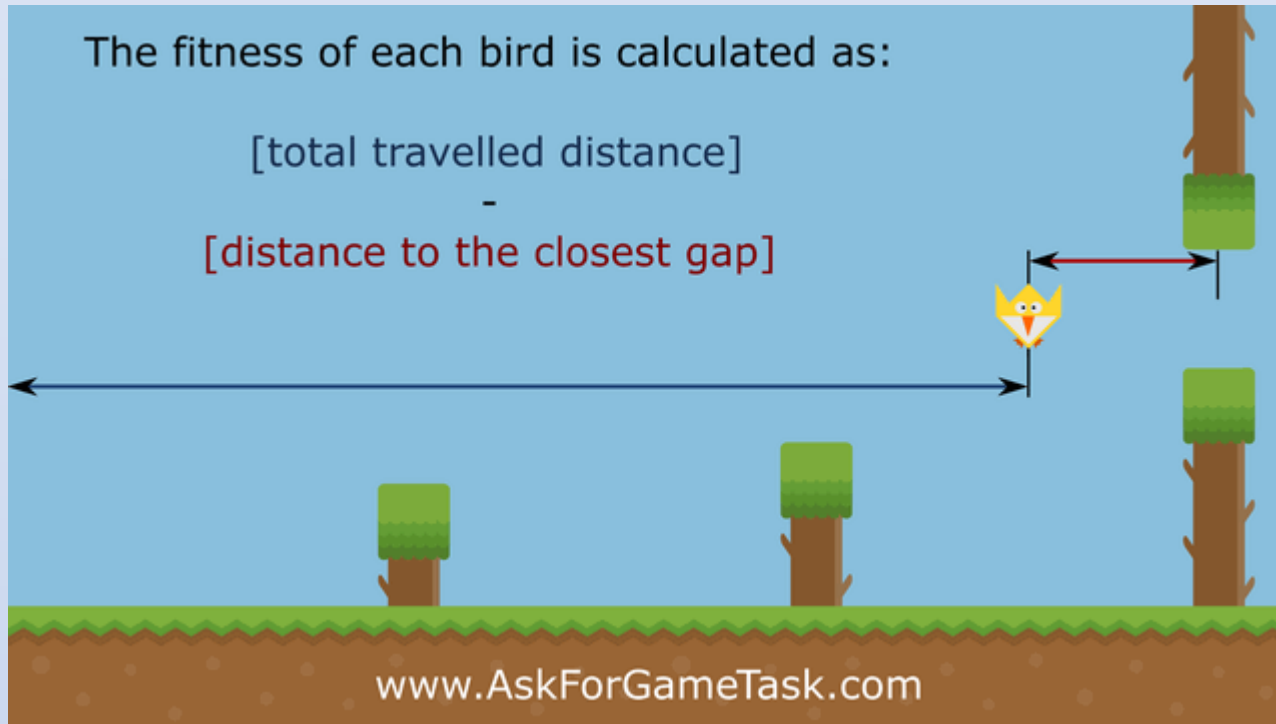


Популяция:
10 птиц, у
каждой своя ИНС
(веса)

Каждая птица (особь популяции) полетела -
узнаем фитнес функцию.

ГА + ИНС

Пример. Игра FlappyBird



Фитнес функция — это разность между общим расстоянием, проделанным птицей, и текущим расстоянием до ближайшего промежутка.

ГА + ИНС

Пример. Игра FlappyBird

Сортируем. Чем больше фитнес функция – тем «лучше» ИНС.

Четыре «лучших» ИНС (четыре победителя) – непосредственно в следующее поколение.

Создается **один** потомок как результат кроссинговера между **двумя** наилучшими победителями.

Создается **три** потомка как результаты кроссинговера двух случайных победителей

Создается **два** потомка как прямые копии двух случайных победителей

К каждому потомку случайные мутации (вариативность).

И цикл!!!

ГА + ИНС

Пример. Игра FlappyBird

Использование ГА для выбора лучшей ИНС
управление птичкой.

Смотри:

<https://habr.com/ru/post/336612/>

Рекомендована ЛІТЕРАТУРА

- Саймон Д. Алгоритмы эволюционной оптимизации. — М.: ДМК Пресс, 2020. — 1002 с.
- Скобцов Ю.А., Федоров Е.Е. Метаэвристики. — Донецк, Изд-во «Ноулидж», 2013. — 426 с.

Посилання

- <https://towardsdatascience.com/gas-and-nns-6a41f1e8146d>
- <https://habr.com/ru/post/246951/>

Контрольні запитання

1. Какие методы применяются для генерации начальной популяции?
2. Какая информация используется при отборе родителей?
3. Какие недостатки имеет "метод рулетки"?
4. Чем отличается ранжирование от пропорционального отбора?
5. Что такое локальный отбор?
6. Опишите метод турнирного отбора.
7. Опишите методы отбора пар для скрещивания.
8. Что такое неявные методы отбора?
9. Опишите двоичную рекомбинацию.

Контрольні запитання

10. Чем отличается многоточечный кроссинговер от классического?
11. Что такое однородный кроссинговер?
12. Чем отличается рекомбинация действительных чисел от классического кроссинговера?
13. Что такое дискретная рекомбинация?
14. Опишите промежуточную рекомбинацию.
15. Чем отличается линейная рекомбинация от промежуточной?
16. Что такое инверсия?

The END
Mod 2. Lec 09