



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Применение генетических алгоритмов для настройки весовых коэффициентов в нейронных сетях

Аспирант 1-го года обучения департамента анализа данных и искусственного интеллекта ФКН НИУ
ВШЭ (Москва)

Махажанов Нуртас

makhazhanovn@gmail.com

Научный руководитель:

д.ф.-м.н., профессор, зав. департамента анализа данных и искусственного интеллекта

Кузнецов Сергей Олегович

23.04.2015

Разделимость множеств

Если RBF-сеть используется для решения задач классификации образов, то основная идея решения обычно состоит в нелинейном преобразовании входных данных в пространстве более высокой размерности.

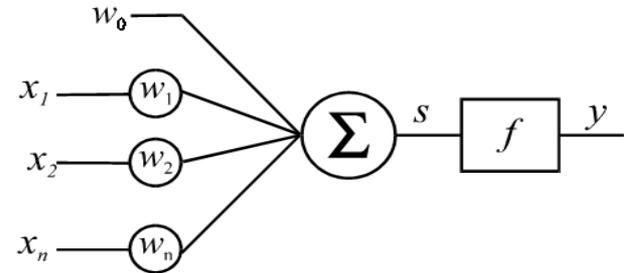
Рассмотрим семейство поверхностей, каждая из которых делит входное пространство на две части. Пусть X – множество, состоящее из N образов (векторов) x_1, x_2, \dots, x_N , каждый из которых принадлежит одному из двух классов X_1 или X_2 . Это бинарное разбиение точек называется *разделимым* по отношению к семейству поверхностей, если в этом семействе существует поверхность, которая отделяет точки класса X_1 от точек класса X_2 . Для каждого образа $x \in X$ определим вектор, состоящий из множества действительных значений функций $\{\varphi_i(x) | i = 1, 2, \dots, m_1\}$, вида

$$\varphi(x) = [\varphi_1(x), \varphi_2(x), \dots, \varphi_{m_1}(x)]^T$$

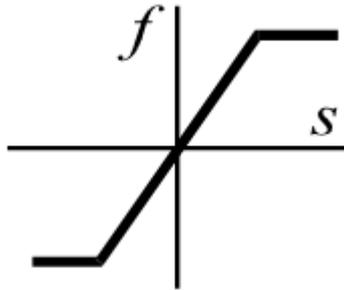
Предположим, что образ x является вектором в m_0 -мерном входном пространстве. Тогда векторная функция $\varphi(x)$ отображает точки m_0 -мерного пространства в новое пространство размерности m_1 . Функции $\varphi_i(x)$ называются *скрытыми*, поскольку они играют роль скрытых элементов нейронных сетей прямого распространения.

Соответственно пространство, образованное множеством скрытых функций $\{\varphi_i(x)\}_{i=1}^{m_1}$, называется *скрытым пространством* или *пространством признаков*.

Сигналы x_i , поступающие на вход нейрона, умножаются на весовые коэффициенты w_i (синаптические веса). Далее они суммируются, и результирующий сигнал, сдвинутый на величину смещения w_0 подается на вход блока, реализующего активационную функцию нейрона.

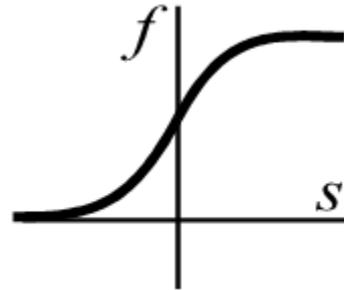


Виды активационных функций:



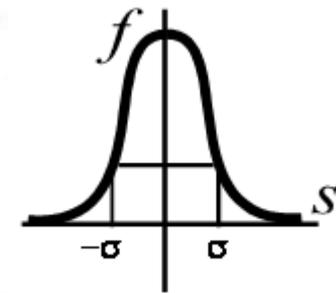
Линейная

$$f(s) = \begin{cases} a, & \text{если } s > a \\ s, & \text{если } -a \leq s \leq a \\ -a, & \text{если } s < -a \end{cases}$$



Сигмоидальная

$$f(s) = \frac{1}{1 + e^{-s}}$$



Радиально-симметричная

$$f(s) = e^{-\frac{(s-x)^2}{\sigma^2}}$$

Разделимость множеств

Дихотомия $\{X_1, X_2\}$ множества X называется φ -разделимой, если существует m_1 -мерный вектор w , для которого можно записать

$$w^T \varphi(x) > 0, x \in X_1,$$

$$w^T \varphi(x) < 0, x \in X_2.$$

Гиперплоскость, задаваемая уравнением

$$w^T \varphi(x) = 0,$$

описывает поверхность в φ -пространстве (т.е. в скрытом пространстве). Обратный образ этой поверхности, т.е.

$$x: w^T \varphi(x) = 0,$$

определяет *разделяющую поверхность (separating surface)* во входном пространстве.

Рассмотрим естественный класс отображений, получаемый при использовании линейной комбинации произведений r -координат вектора входного образа.

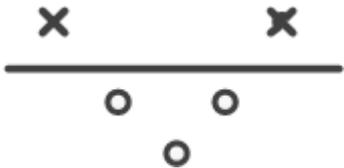
Разделяющие поверхности, соответствующие такому отображению, называются *рациональными многообразиями r -го порядка*.

Рациональное многообразие r -го порядка в пространстве размерности m_0 описывается однородным уравнением r -го порядка в координатах входного вектора x :

$$\sum_{0 \leq i_1 \leq i_2 \leq \dots \leq i_r \leq m_0} a_{i_1, i_2, \dots, i_r} x_{i_1} x_{i_2} \dots x_{i_r} = 0$$

где x_i - i -й компонент входного вектора x . Произведение элементов x_i вектора x r -го порядка (т.е. $x_{i_1} x_{i_2} \dots x_{i_r}$) называется *одночленом* (monomial). Для входного пространства размерности m_0 сумма включает $\frac{(m_0-r)!}{m_0!r!}$ одночленов.

Примеры φ -разделимых дихотомий для различных множеств из пяти точек в двумерном пространстве:



линейно-разделимая
дихотомия



сферически разделимая
дихотомия

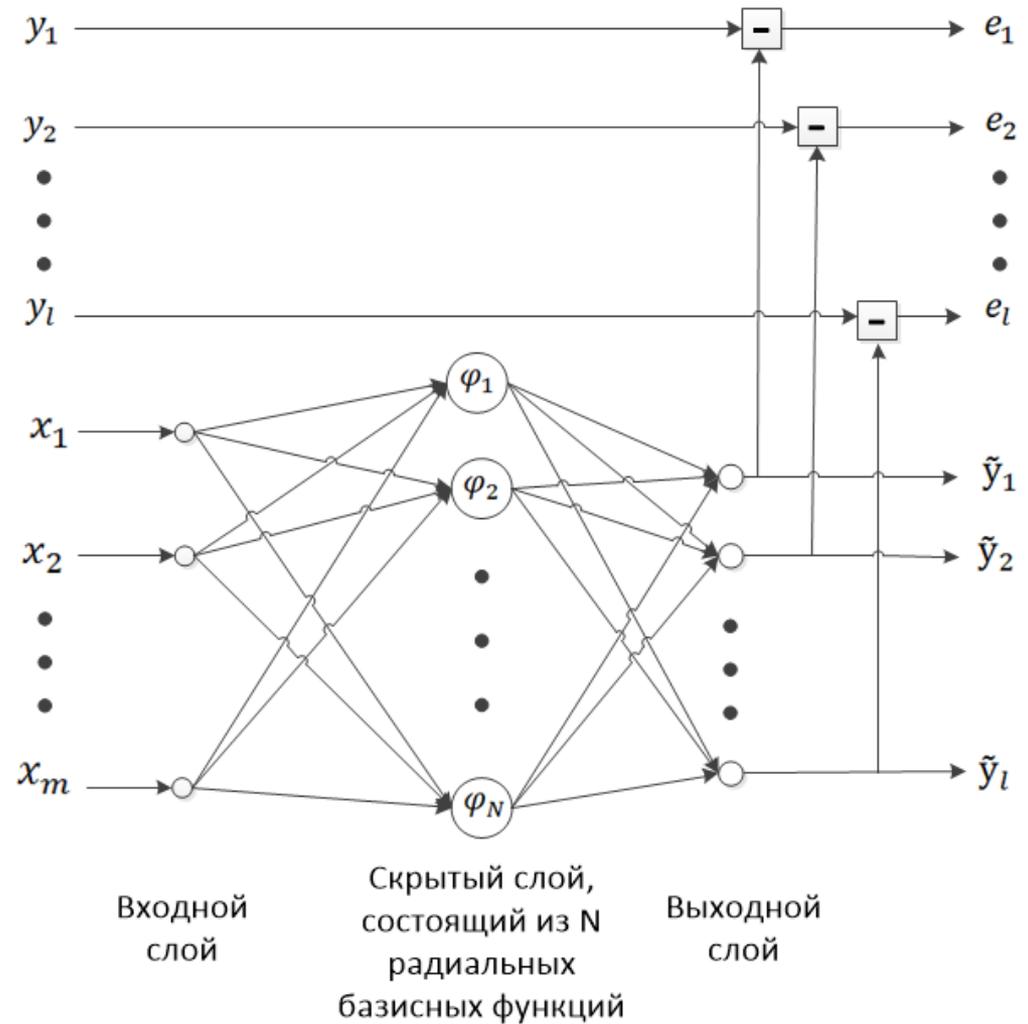


квадратично-разделимая
дихотомия



Нейронная сеть на основе радиальных базисных функций

Обучение нейронной сети с учителем предполагает, что для каждого входного вектора из обучающего множества существует требуемое значение выходного вектора, называемого целевым. Эти вектора образуют обучающую пару. Веса сети изменяют до тех пор, пока для каждого входного вектора не будет получен приемлемый уровень отклонения выходного вектора от целевого.



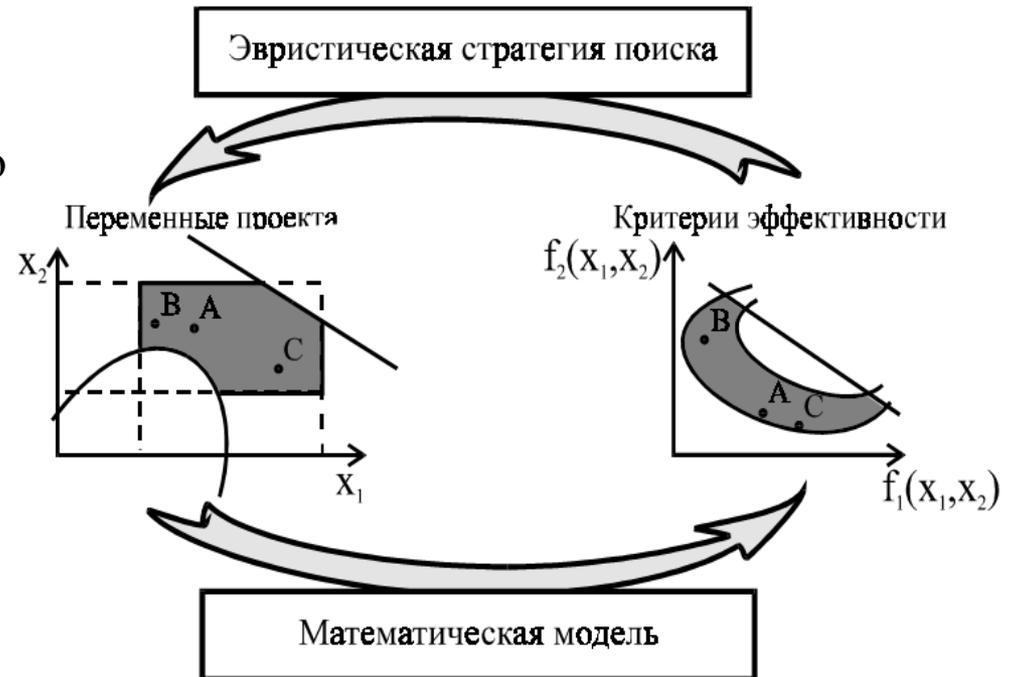
Цель в оптимизации с помощью Генетических алгоритмов (ГА) состоит в том, чтобы найти лучшее возможное решение или решения задачи по одному или нескольким критериям. Чтобы реализовать генетический алгоритм нужно сначала выбрать подходящую структуру для представления этих решений. В постановке задачи поиска, экземпляр этой структуры данных представляет точку в пространстве поиска всех возможных решений.

Структура данных генетического алгоритма состоит из одного или большего количества хромосом (обычно из одной). Как правило, хромосома - это битовая строка, так что термин строка часто заменяет понятие "хромосома". В принципе, ГА не ограничены бинарным представлением. Известны другие реализации, построенные исключительно на векторах вещественных чисел (L. Davis, 1991b; Eshelman и Schaffer, 1993; Goldberg, 1991a, 1991b). Несмотря на то, что для многих реальных задач, видимо, больше подходят строки переменной длины, в настоящее время структуры фиксированной длины наиболее распространены и изучены. Пока и мы ограничимся только структурами, которые являются одиночными строками по n бит.

Модель генетических алгоритмов

В пространстве всех возможных решений, генетические алгоритмы позволяют отобрать такие комбинации параметров, при которых минимизируется функция ошибки.

В определенном смысле можно утверждать, что “вектор переменных проектирования” играет в технике такую же роль, что и категория “генотип” в биологии. Подобно тому, как в Природе скрещивание организмов осуществляется на генетическом уровне, в процедуре оптимизации координаты новых пробных точек получается как результат манипулирования координатами старых.



Алгоритм синтеза RBF-сети включает выбор размера скрытого слоя равным количеству тренировочных шаблонов. Но в случае, если шаблонов очень много, то работать с такой сетью будет сложно. На практике очень трудно сказать априори, какие именно эпизоды взаимодействия с объектом могут стать источником хороших шаблонов для построения нейросетевой модели объекта.

Для того чтобы выяснить, какие шаблоны являются действительно ключевыми в понимании динамики объекта, скомбинируем методику настройки синаптических весов RBF-сети с Генетическими алгоритмами (ГА).

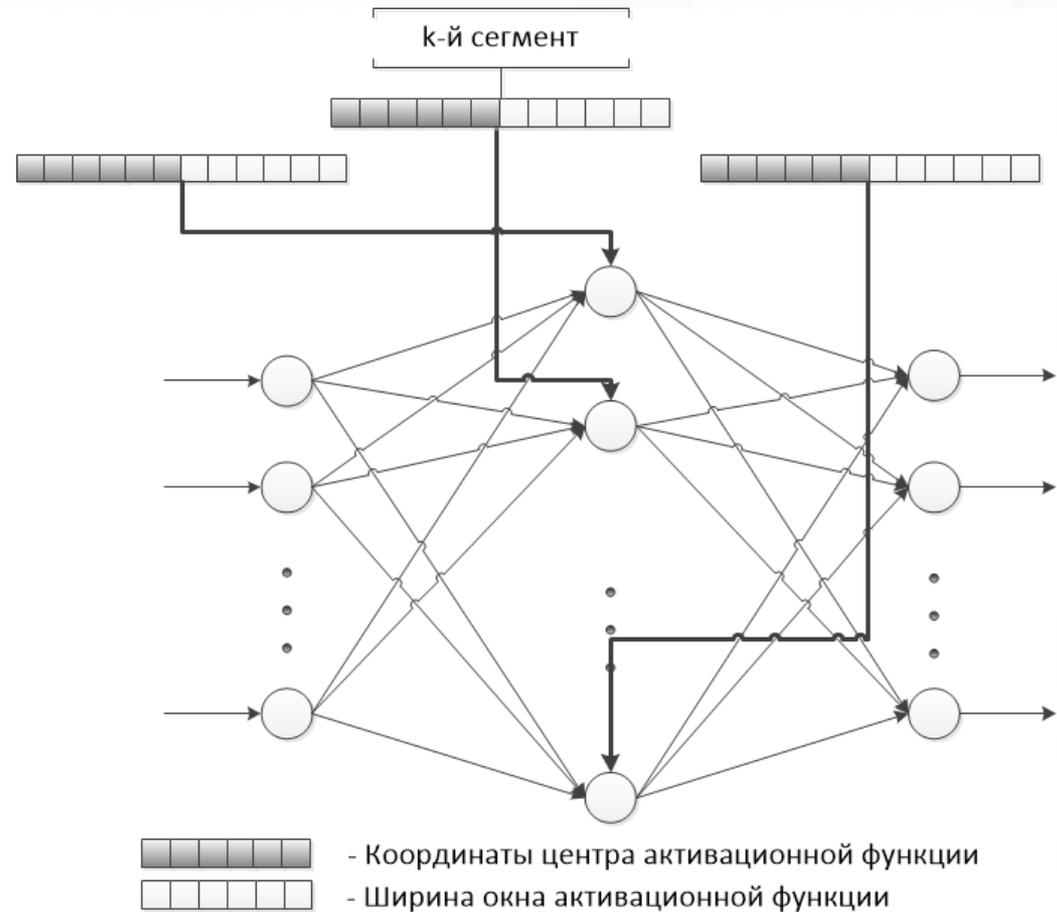
Будем выбирать центры радиальных базисных функций на основе ГА.

Этот подход состоит из двух этапов:

1. *Этап обучения на основе ГА.* Его целью является оценка подходящих положений центров радиальных базисных функций скрытого слоя.
2. *Этап обучения с учителем.* На этом этапе создание сети завершается оценкой линейных весов выходного слоя.

На каждой итерации поиска ГА самостоятельно выбирает, в каких точках пространства входных сигналов сети разметить центры активационных функций нейронов скрытого слоя, и назначает для каждой из них ширину окна.

При декодировании хромосомы в вектор переменных одновременно с величиной настроечных параметров конкретизируется сама структура скрытого слоя сети.

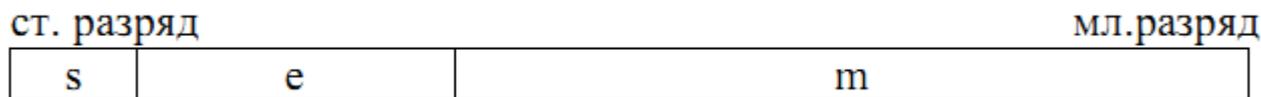


Числа, соответствующие каждой межсинаптической связи в скрытом слое нейронной сети кодируются в двоичный код с плавающей запятой. При использовании десятичной системы счисления с плавающей запятой записываются в виде $m \cdot 10^e$, где m - мантисса, а e – экспонента.

Значение мантиссы определяется как: $m = a_0 \cdot 2^0 + a_{-1} \cdot 2^{-1} + \dots + a_{-n} \cdot 2^{-n}$

Значение экспоненты определяется как значение экспоненты числа в двоичном коде со смещением $2^{l/2} - 1$

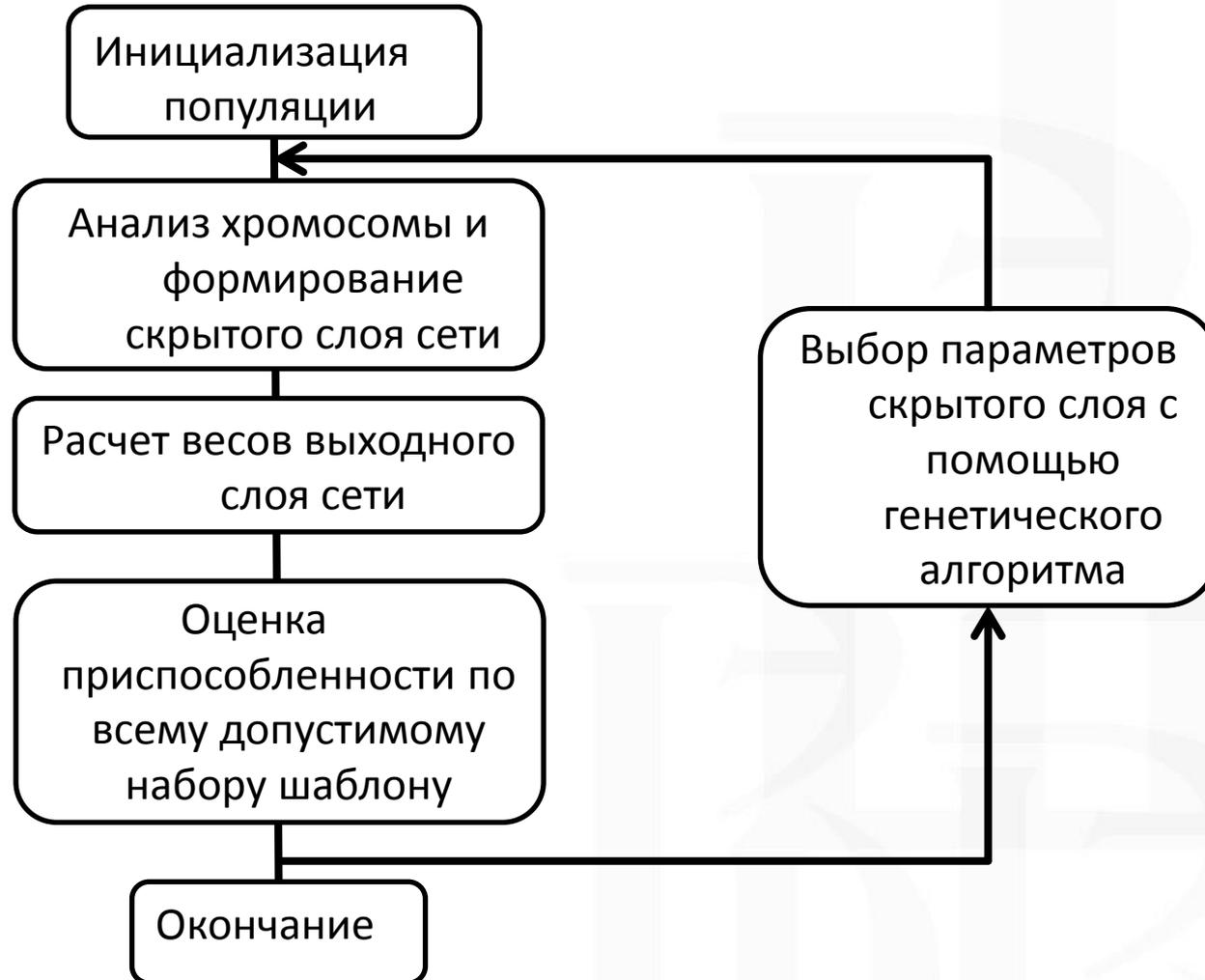
Для представления двоичных чисел с плавающей запятой в настоящее время стандартным в практических приложениях является формат IEEE-754. Формат представления числа с плавающей запятой в этом формате описывается следующим образом:



где s – знаковый разряд, e – экспонента, m – мантисса.



Процедура тренировки RBF-сети



Различия между гаплоидной и диплоидной моделями ГА

Обычно одной точке из пространства переменных соответствует одна точка из пространства критериев. Этот примитивный уровень сложности представлений соответствует способу хранения информации у гаплоидных организмов.

У диплоидных организмов особь определяется двумя точками одновременно (отцовская и материнская хромосомы), хотя этой паре по-прежнему соответствует одна точка из пространства фенотипов. В процессе выработки своей половой клетки такая двухточечная особь генерирует третью точку, которая содержит в себе частичную информацию об обеих родительских хромосомах.

Главным преимуществом диплоидных моделей перед гаплоидными состоит в их генетическом разнообразии. Для гаплоидных популяций *фенотипическое вырождение* особей неизбежно означает и *генотипическое вырождение*. В то же время, для диплоидных особей это правило не является обязательным. Т.е. фенотипическое вырождение популяции не свидетельствует об утрате генетического разнообразия.

Доминантные и рецессивные признаки

Введем четырехбуквенный алфавит

$$\{R, r, D, d\},$$

в котором R и r обозначают *рецессивную 1* и *рецессивный 0*, а D и d – *доминантную 1* и *доминантный 0*.

Сформулируем правила полного доминирования генов:

		Гены более приспособленной особи			
		D	d	R	r
Гены менее приспособленной особи	D	1	0	1	1
	d	1	0	0	0
	R	1	0	1	1
	r	1	0	0	0

Триада генетических алгоритмов: кроссинговер

Родитель A
(более приспособленный)

D R r R D D d R

r r R d r D R R

Гаметы A

D r R d r D R R

~~r R r R D D d R~~

Родитель B
(менее приспособленный)

d d R D D d d R

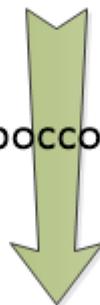
R r r D d d R r

Гаметы B

~~d r r D d d R r~~

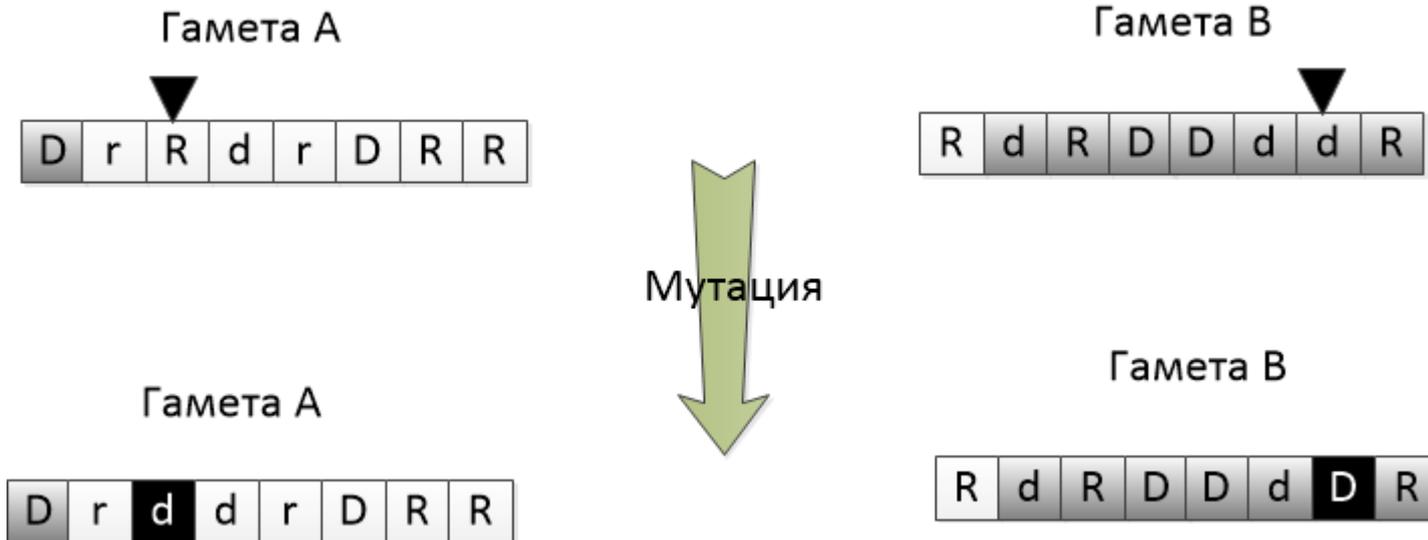
R d R D D d d R

Кроссовер



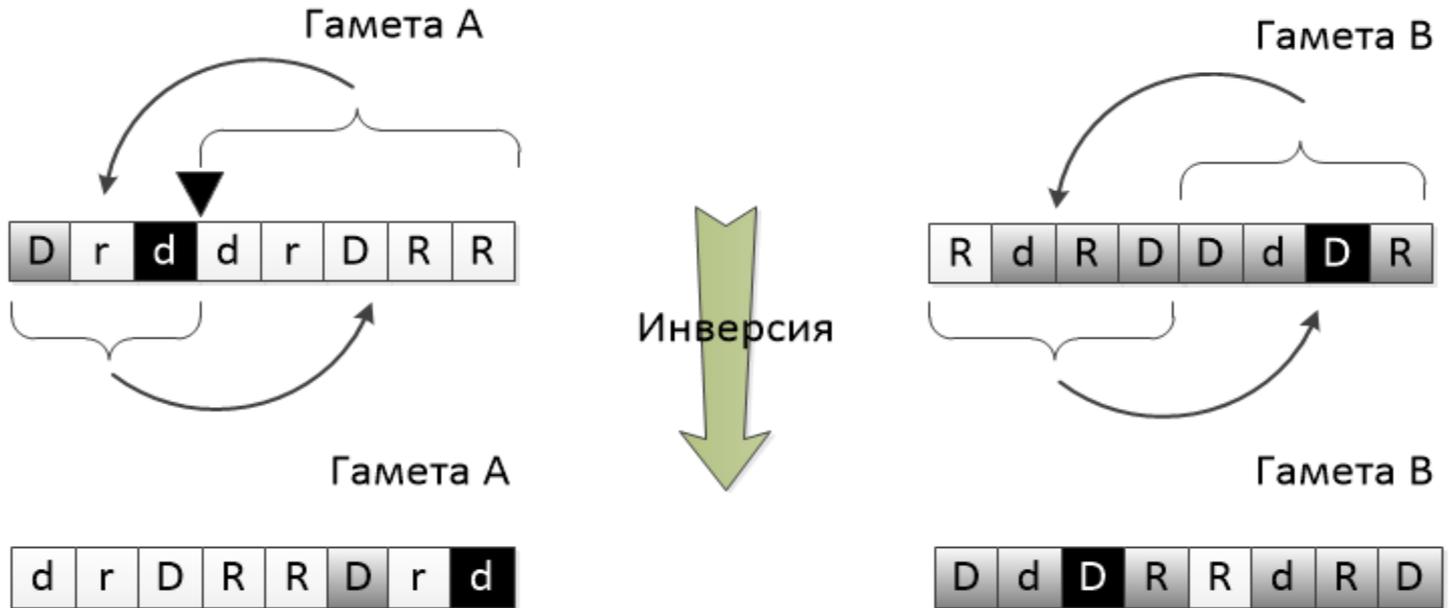


Триада генетических алгоритмов: мутация





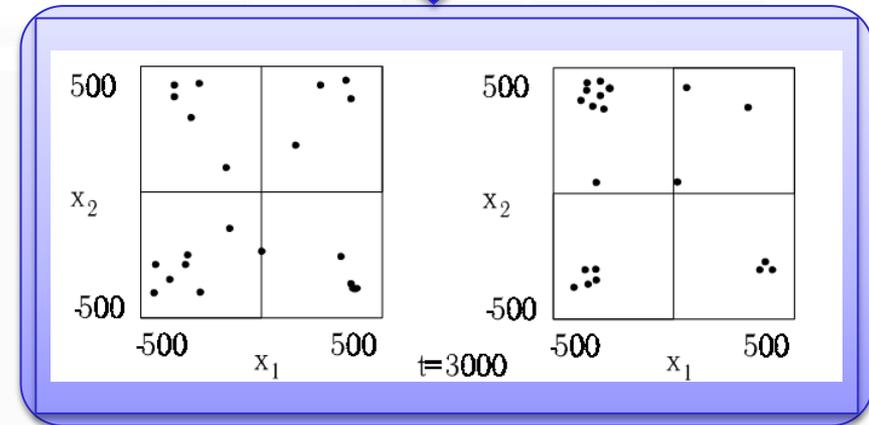
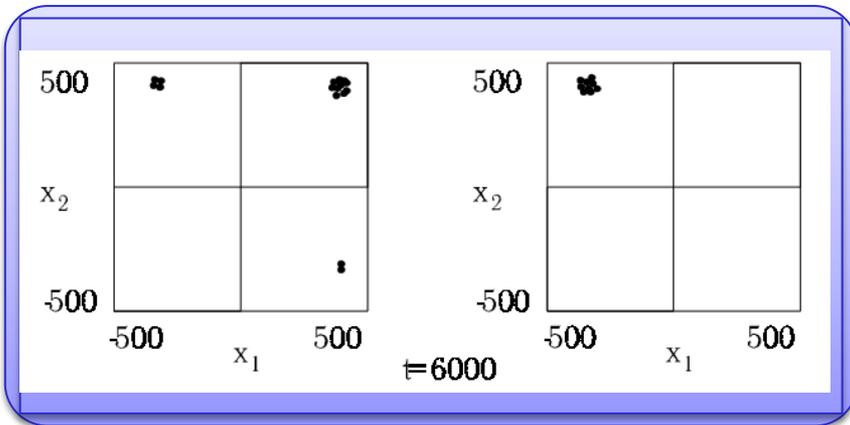
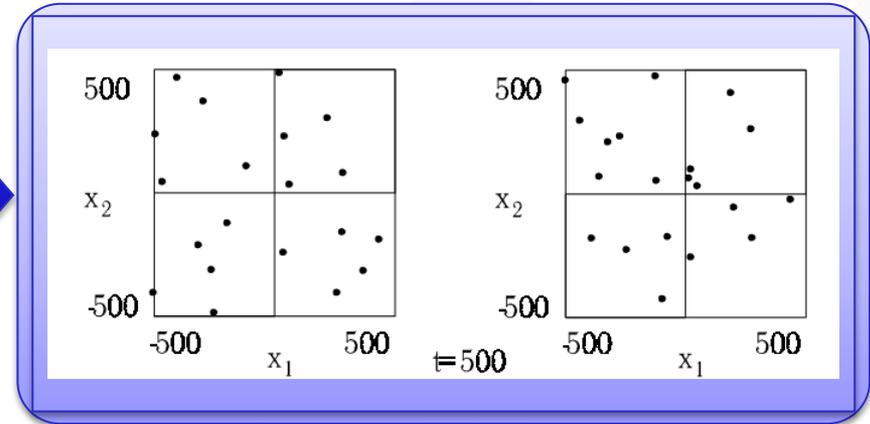
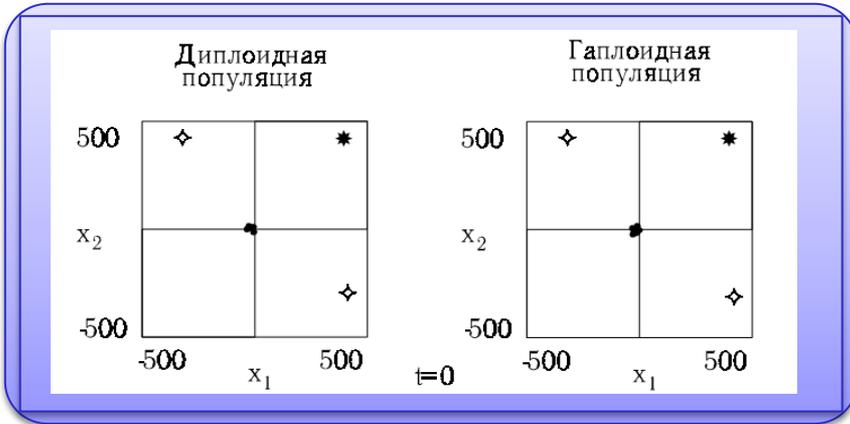
Триада генетических алгоритмов: инверсия



Зигота потомка



Поиск экстремума тестовой функции предварительно вырожденными диплоидной и гаплоидной популяциями.



★ - глобальный экстремум

◇ - локальный экстремум

● - особь популяции

Селекция нейронных сетей

Функция приспособленности особи вычисляется на основе сравнения ее прогнозов с истинными значениями.

На интервале в 50 шагов, предшествующих текущему моменту, особь делает прогноз, и на каждом шагу считается ошибка прогноза. Просуммировав все ошибки, мы получаем общую ошибку прогноза каждой нейронной сети. Эта ошибка и будет служить мерой приспособленности данной особи к окружающей среде.

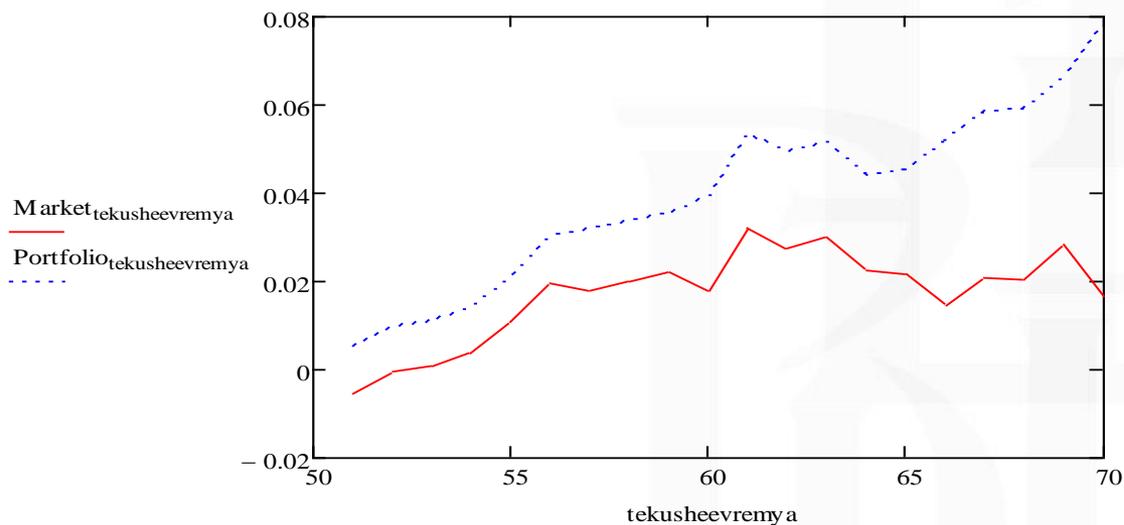
Селекция – это выбор тех хромосом, которые будут участвовать в создании потомков для следующей популяции. Такой выбор производится согласно принципу естественного отбора, по которому наибольшие шансы на участие в создании новых особей имеют хромосомы с наибольшими значениями функции приспособленности. Существуют различные методы селекции. Далее среди всех особей популяции проводится селекция по принципу точности прогноза.



Сравнение портфеля на основе метода ИНС и ГА с рыночным портфелем

Применение нейронных сетей целесообразно, если:

1. накоплены достаточные объемы данных о предыдущем поведении системы
2. не существует традиционных методов или алгоритмов, которые удовлетворительно решают проблему
3. данные частично искажены, частично противоречивы или не полны и поэтому традиционные методы выдают неудовлетворительный результат.



Мозг – орган, экспрессирующий наибольшее число генов организме. На молекулярном уровне специфичность каждой из клеток организма создается составом белков, из которых она построена и которые обеспечивают ее функции. Эти белки синтезируются за счет активности генов в составе ДНК, содержащейся в ядре клетки - в геноме организма. Синтез белка посредством считывания информации с гена в виде молекулы матричной РНК (мРНК) и ее трансляции в белковую молекулу называется экспрессией гена.

Ясно, что в эволюции мозга, как и других органов тела, критическую роль играли регуляторные гены, определяющие процессы эмбрионального развития. Приобретение ими функций в нервной системе должно было происходить под контролем естественного отбора, дающие увеличение преимущества в выживании и/или размножении. Эти функции могли осуществляться на двух фазах эволюционного цикла. Одна из них – формирования инстинктивного «врожденного» поведения. Но, в отличие от других соматических органов, в мозге многие из этих генов вновь активируются и после завершения созревания – в ситуациях новизны и обучения.

Таким образом, в отношении мозга две фазы эволюционного цикла – созревание и адаптивные модификации функциональных систем, обеспечивающих выживание, оказываются тесно связанными на уровне механизмов регуляции экспрессии генов. По сути дела в мозге процессы морфогенеза и развития никогда не прекращаются, а лишь переходят под контроль когнитивных процессов.

ГА можно использовать для определения некоторых общих свойств нейронной сети, а не для определения числовых характеристик конкретных связей. Например, в генотипе особей можно шифровать:

- матожидание и дисперсию ветвистости нейронов на каждом выделенном сегменте нейронной сети,
- показатели input-а на которые будет направлено “внимание” нейронной сети,
- информацию о самоорганизации сети.

А затем можно строить нейронные сети со случайными связями между нейронами, но так, чтобы соблюдались общие свойства сети, зашифрованные в генотипе особи.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Спасибо за внимание!

Махажанов Нургас

Аспирант 1-го года обучения департамента анализа данных и искусственного интеллекта ФКН НИУ ВШЭ

makhazhanovn@gmail.com