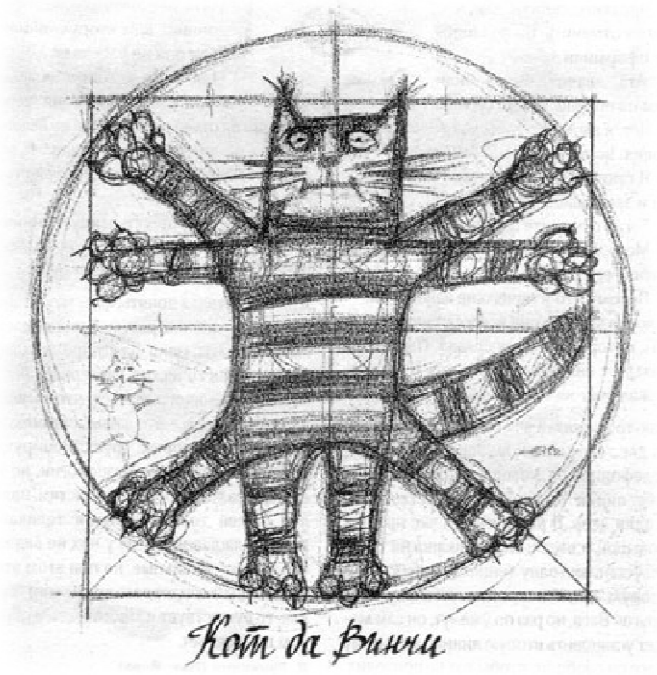
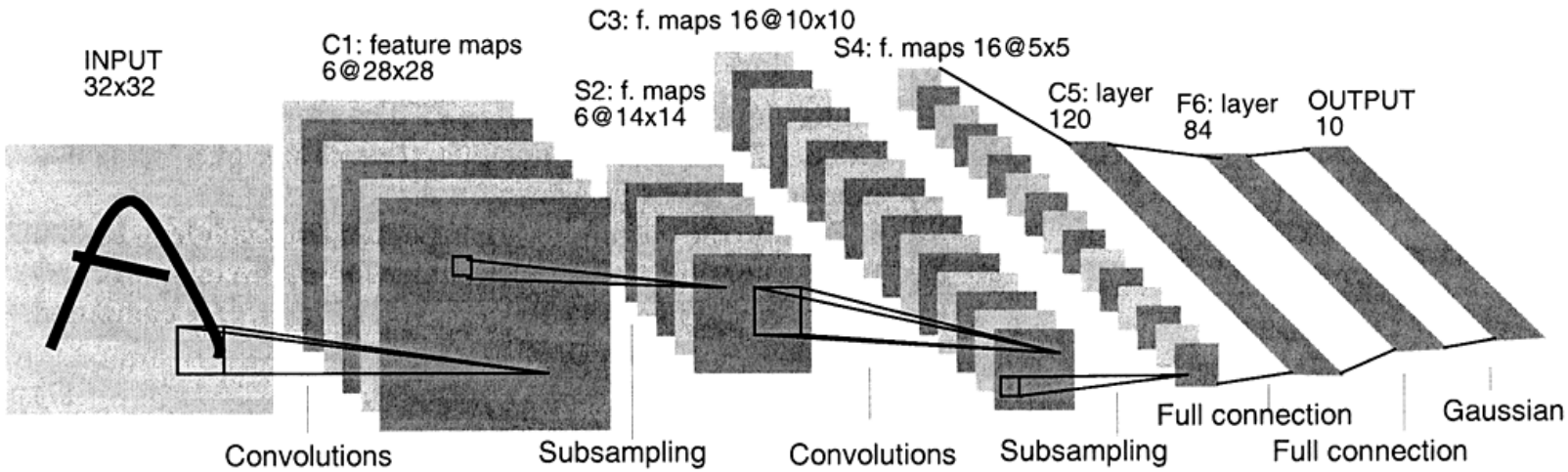


Нейроэволюционное преобразование пространства признаков в задаче нейросетевой классификации

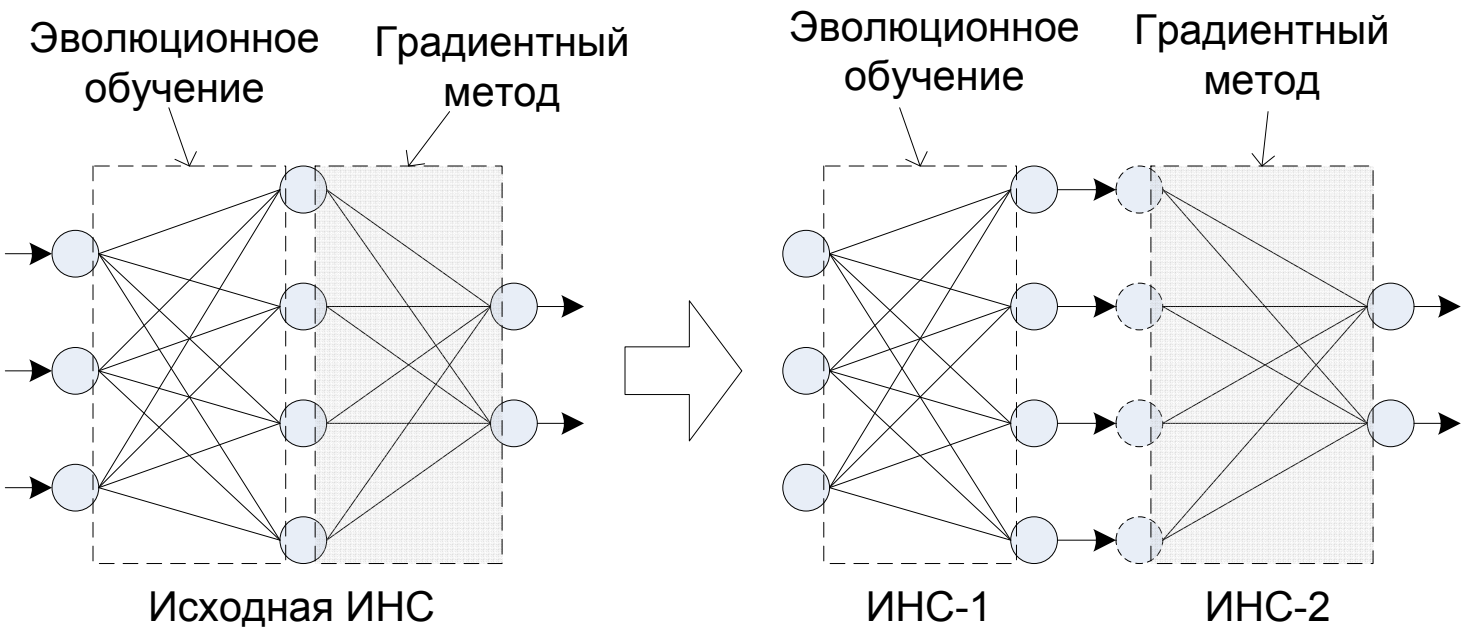
Цой Ю.Р.

Преобразование пространства признаков

- Понижение размерности (отбор признаков, метод главных компонент, разложение по базису из ортогональных функций)
- Повышение размерности (сверточные сети, ИНС, глубокое обучение (deep learning))



Комбинированное обучение



ИНС не обязательно является многослойной!

Предыдущие результаты (КИИ-2010)

Целевая функция: ошибка обучения ИНС-2.
Сложность подбора параметров. Большое время.

Задача	Лучший полученный результат	Результаты из [Prechelt, 1994]	Комментарии
cancer1	2,07 (0,73)	1,38 (0,49)	t_1=10, t_2=100, alpha = 2,5
card1	10,17 (0,63)	14,05 (1,03)	t_1=10, t_2=100, alpha = 2,0
diabetes1	21,41 (0,16)	24,10 (1,91)	t_1=10, t_2=500, alpha = 2,0
glass1	26,41 (0)	32,70 (5,34)	t_1=10, t_2=400, alpha = 2,0
heart1	17,74 (1,02)	19,72 (0,96)	t_1=20, t_2=100, alpha = 2,0
horse1	34,40 (3,84)	29,19 (2,62)	t_1=30, t_2=100, alpha = 2,0

Другая целевая функция?

Корреляционный критерий

$$f = \frac{2\sum_{i=1}^{N}\sum_{j>i}^{N} R_{Y^i,Y^j}}{N(N-1)}, \qquad R_{Y^i,Y^j} = \frac{\sum_k (y_k^i - \overline{Y}^i)(y_k^j - \overline{Y}^j)}{N\sigma_{Y^i}\sigma_{Y^j}},$$

Два варианта:

- $f \rightarrow \max$
- $f \rightarrow \min$

Снижение размерности

Повышение размерности

Тестовые задачи

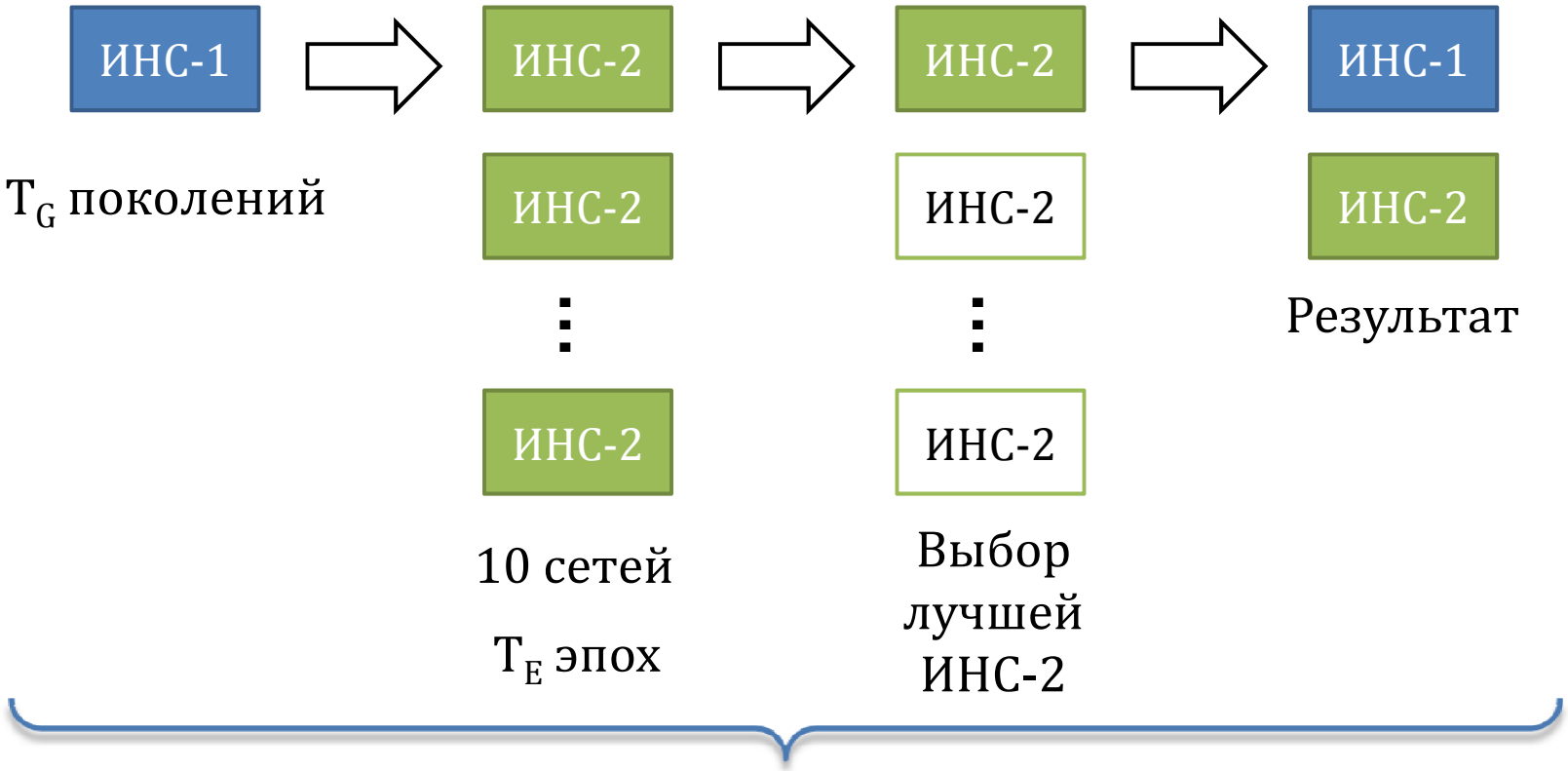
Тестовый набор: **Proben1**

Задача	Количество примеров в обучающей и тестовой выборках	Число входных признаков	Количество классов
cancer1	699	9	2
card1	690	51	2
diabetes1	768	8	2
glass1	214	9	6
heart1	920	35	2
horse1	364	58	3

RPROP: Encog (http://www.heatonresearch.com/encog)

Комбинированное обучение: Mental Alchemy (http://code.google.com/p/mentalalchemy/)

Комбинированное обучение. Настройка



10 запусков

Рассматриваются результаты лучшего запуска

Результаты

Сложность подбора параметров.
Требует меньше времени для обучения.

Задача	Мин. корелляции	Макс. корелляции	Мин. ошибки	Стандартное обучение ИНС
cancer1	1,03 (0,53)	1,26 (0,45)	2,07 (0,73)	1,38 (0,49)
card1	11,28 (0,30)	12,21 (0,39)	10,17 (0,63)	14,05 (1,03)
diabetes1	22,55 (0,95)	22,24 (0,65)	21,51 (0,43)	24,10 (1,91)
glass1	25,66 (0,97)	27,17 (0,97)	26,41 (0)	32,70 (5,34)
heart1	17,04 (0,34)	18,43 (0,42)	17,74 (1,02)	19,72 (0,96)
horse1	24,51 (0,53)	28,35 (2,25)	34,40 (3,84)	29,19 (2,62)

Обсуждение

Промежуточные выводы:

- Большое количество входных признаков: Линейные функции активации и максимизация корреляции.
- Малое количество входных признаков: Сигмоидная функции активации и минимизация корреляции.
- Сложность настройки параметров.
- Обучение с корреляционным критерием быстрее, чем при минимизации ошибки ИНС-2.
- Классические комитетные методы (Bagging, AdaBoost.M1) не принесли улучшений.

Анализ корреляционного критерия

Ковариация между двумя векторами

$$Cov(\mathbf{Y_1}, \mathbf{Y_2}) = \frac{1}{N}(\mathbf{Y_1} - \overline{Y_1}\mathbb{1})^{\mathsf{T}}(\mathbf{Y_2} - \overline{Y_2}\mathbb{1}),$$

$$NCov(\mathbf{Y_1}, \mathbf{Y_2}) = \mathbf{Y_1}^{\mathsf{T}}\mathbf{Y_2} + r,$$

$$r = -N\overline{Y_1 Y_2}.$$

Многие функции активации (сигмоидная, радиальная и др.) принимают значения в диапазоне [0; 1]!

Объем параллелограмма, построенного на $\mathbf{Y}_i$, меньше максимально возможного.

Новый критерий!

- Используется матрица Грамма G для нормализованных по L2 выходных сигналов ИНС-1.
- Штраф за наличие нулевых диагональных элементов в G.

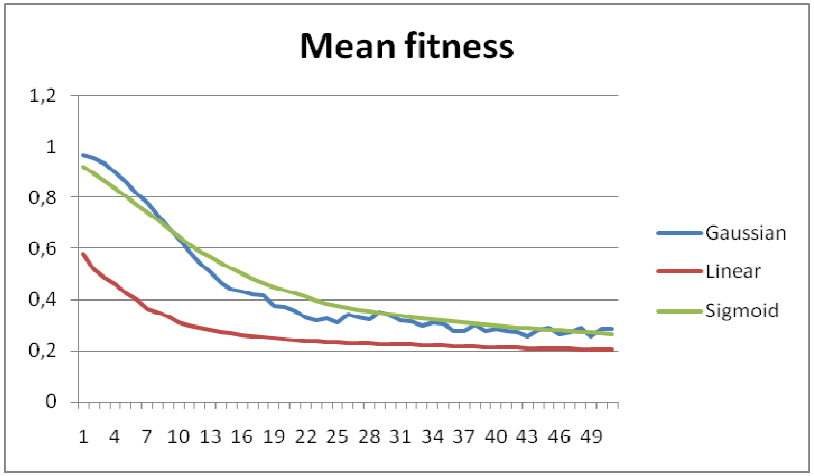
$$f = \frac{2 \sum\limits_{i,j>i} G(i,j)}{\alpha n(\alpha n - 1)} + \pi \to min,$$

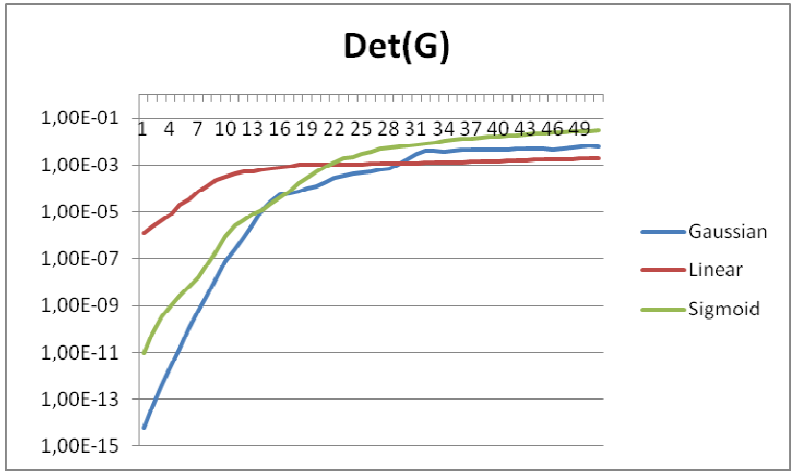
Результаты

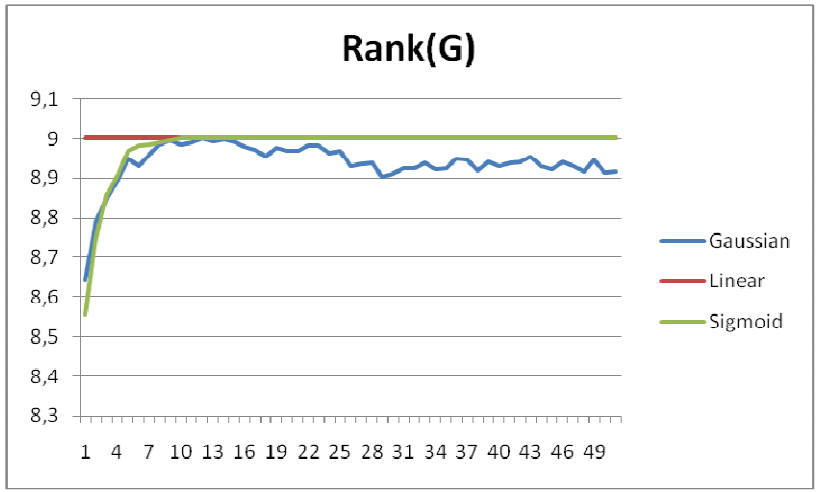
Меньше чувствительность к настройкам параметров.

	Новый критерий	Мин. корелляции	Макс. корелляции	Стандартное обучение ИНС [5]
cancer1	0,92 (0,47)	1,03 (0,53)	1,26 (0,45)	1,38 (0,49)
card1	11,98 (0,74)	11,28 (0,30)	12,21 (0,39)	14,05 (1,03)
diabetes1	22,14 (0,56)	22,55 (0,95)	22,24 (0,65)	24,10 (1,91)
glass1	27,92 (1,19)	25,66 (0,97)	27,17 (0,97)	32,70 (5,34)
heart1	17,17 (0,55)	17,04 (0,34)	18,43 (0,42)	19,72 (0,96)
horse1	24,73 (2,21)	24,51 (0,53)	28,35 (2,25)	29,19 (2,62)

Анализ эволюции ИНС-1. Cancer1.



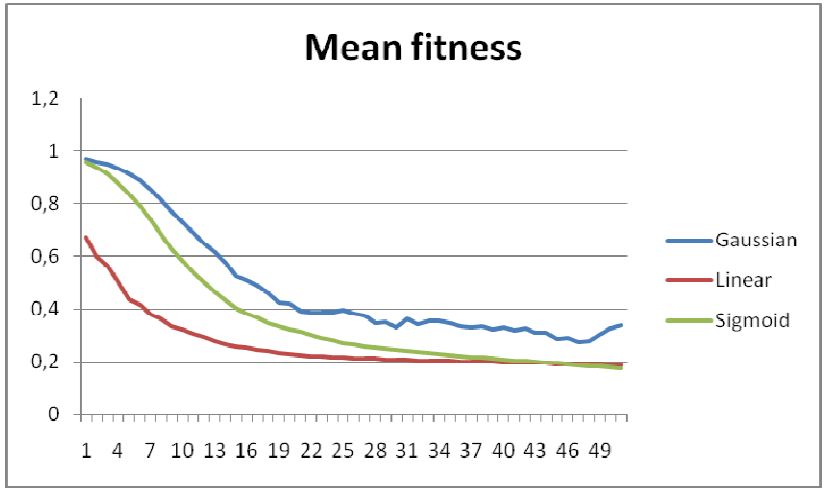


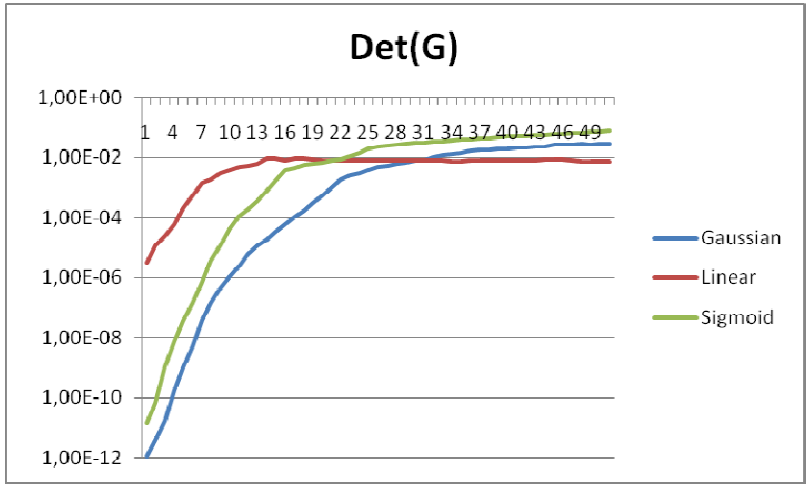


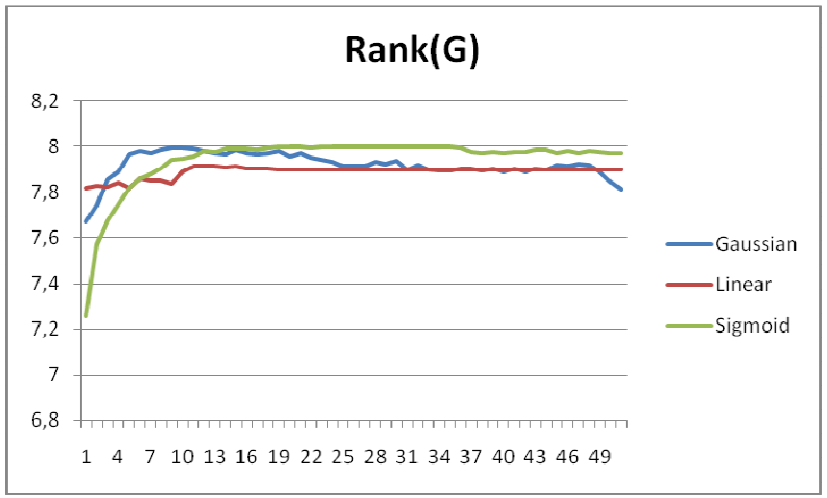
	Gaussian	Linear	Sigmoid
a = 1	2.30 (0.61)	2.18 (0.45)	2.24 (0.57)

	Gaussian	Linear	Sigmoid
a = 1	1.72 (0.00)	1.72 (0.00)	1.67 (0.18)

Анализ эволюции ИНС-1. Diabetes1.





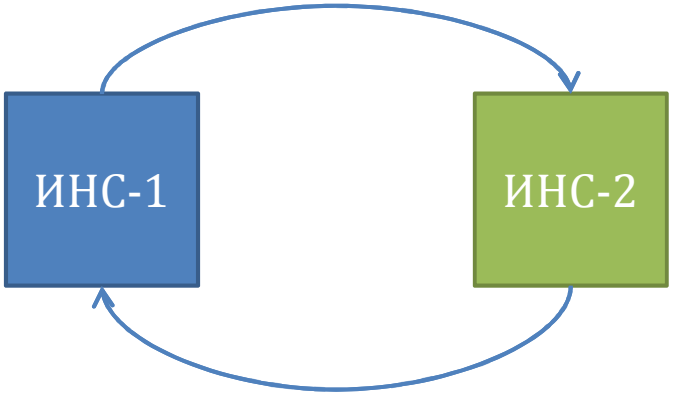


	Gaussian	Linear	Sigmoid
a = 1	27.76 (2.21)	25.68 (2.13)	30.99 (4.35)

	Gaussian	Linear	Sigmoid
a = 1	24.21 (0.93)	22.86 (5.18)	24.84 (0.70)

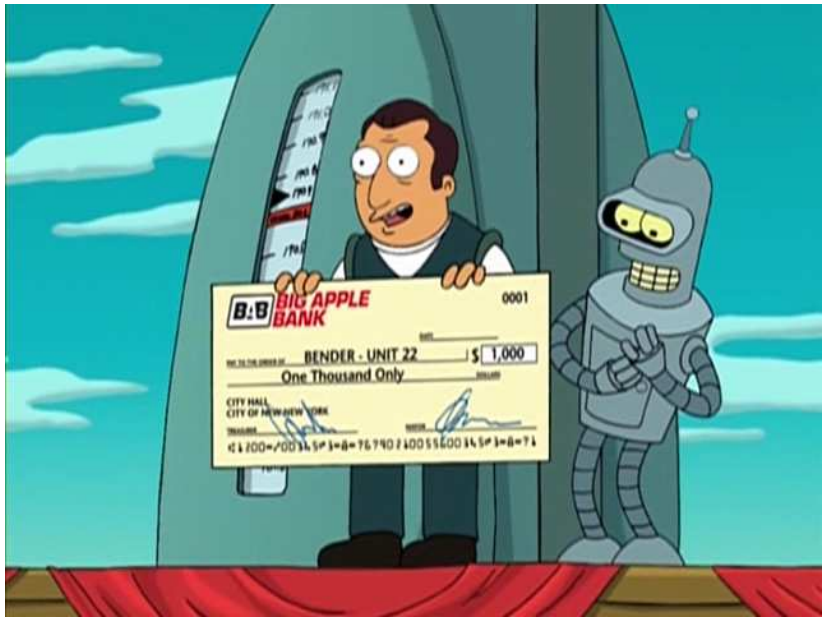
Дальнейшие исследования

- Почему предлагаемый способ работает? Наиболее «простой» способ повышения размерности?
- «Model Selection». Регуляризация по …?
- Автоматический подбор функций активации и размера выходного слоя ИНС-1.
- Инкрементное обучение.



Благодарности

Работа выполнена при поддержке РФФИ (проекты № 09-08-00309-а, 11-07-00027-а)



Спасибо за внимание!

Нейроэволюционное преобразование пространства признаков в задаче нейросетевой классификации

Цой Ю.Р.