

Усовершенствованный искусственный интеллект

Александр Горбань

Посвящается памяти Р.Г. Хлебопроса



Plan

- Self-esteem and error functions for neural networks;
- Ensemble learning and distribution of problems between local experts;
- Adaptive coding and self-supervising learning;
- Organisation of social networks of neural agents and non-iterative knowledge transfer.

Три пришествия нейронных сетей

- Розенблатт и Уидроу (конец 1950х - начало 1960х);
- Хопфилд и Хинтон с соавторами (1980е);
- GOOGLE, Amazon и другие гиганты (2000е - сейчас).

Технические возможности компьютеров накапливались монотонно.

Интеллектуальные прорывы в области нейронных сетей опережали технические возможности, а потом замедлялись до появления нового поля возможностей.

Терпеливые стали классиками!

Причины третьего пришествия

- Компьютеры стали достаточно мощными;
- Информационные гиганты (GOOGLE, AMAZON, FACEBOOK, MICROSOFT) в достаточной мере развили бизнес сетевых услуг;
- 10 марта 2000 года, индекс NASDAQ упал более чем в полтора раза при закрытии;
- Необходимость диверсификации дот-комов;
- Красивые технологические находки, оформленные в 2000-е. “This is deep learning’s Cambrian explosion.” Deep learning – много слоев нейронов друг за другом.

Кризис курса акций Amazon 2000-2001



Несколько базовых идей.

Коннекционизм

- Однородность системы (элементы одинаковы и чрезвычайно просты, все определяется структурой связей);
- Каждая связь представляется как совсем простой элемент, служащий для передачи сигнала.
- Надежные системы из ненадежных элементов и «аналоговый ренессанс» использование простых аналоговых элементов;
- «Голографические» системы при разрушении случайно выбранной части система сохраняет свои свойства.

GOOGLE, Amazon продают

- Услуги нейронных сетей по распознаванию;
- Услуги стартапам по выращиванию нейронных сетей: «Google Cloud's AI provides **modern machine learning services, with pre-trained models** and a service to generate **your own tailored models.**»
- Услуги, использующие нейронные сети: «Major Google applications use Cloud machine learning, including Photos (image search), the Google app (voice search), Translate, and Inbox (Smart Reply).»

Профессии нейронных сетей

Распознавание образов:

- изображений, человеческих лиц, букв и иероглифов, отпечатков пальцев в криминалистике, речи, сигналов радара и сонара,
- элементарных частиц и происходящих с ними физических процессов (эксперименты на ускорителях или наблюдение за космическими лучами),
- заболеваний по симптомам (в медицине),
- местностей, где следует искать полезные ископаемые (в геологии, по косвенным признакам),
- признаков опасности в системах безопасности,
- свойств химических соединений по структуре.

Профессии нейронных сетей

Управление в реальном времени:

- автомобилями,
- самолётами и ракетами,
- станками и машинами,
- технологическими процессами непрерывного производства (в энергетике, металлургии и др.).

Профессии нейронных сетей

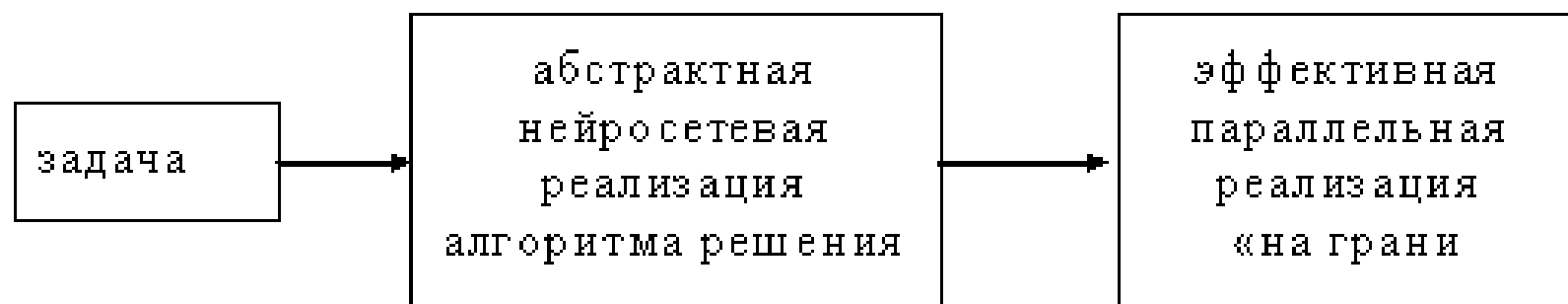
Прогнозирование в реальном времени:

- погоды,
- курса акций (и других финансовых показателей),
- исхода лечения,
- политических событий (результатов выборов, международных отношений и др.),
- поведения противника (реального или потенциального) в военном конфликте и в экономической конкуренции,
- устойчивости супружеских отношений.

Профессии нейронных сетей

- Оптимизация — поиск наилучших вариантов: при конструировании технических устройств, при выборе экономической стратегии, при лечении, при подборе команды.
- Протезирование («умные протезы») и усиление естественных функций.
- Перевод текстов на естественных языках.

Миссия нейронных сетей в эволюции компьютеров



Путь к эф ф е к т и в н о м у п а р а л л е л и з м у.

Элементы нейронных сетей

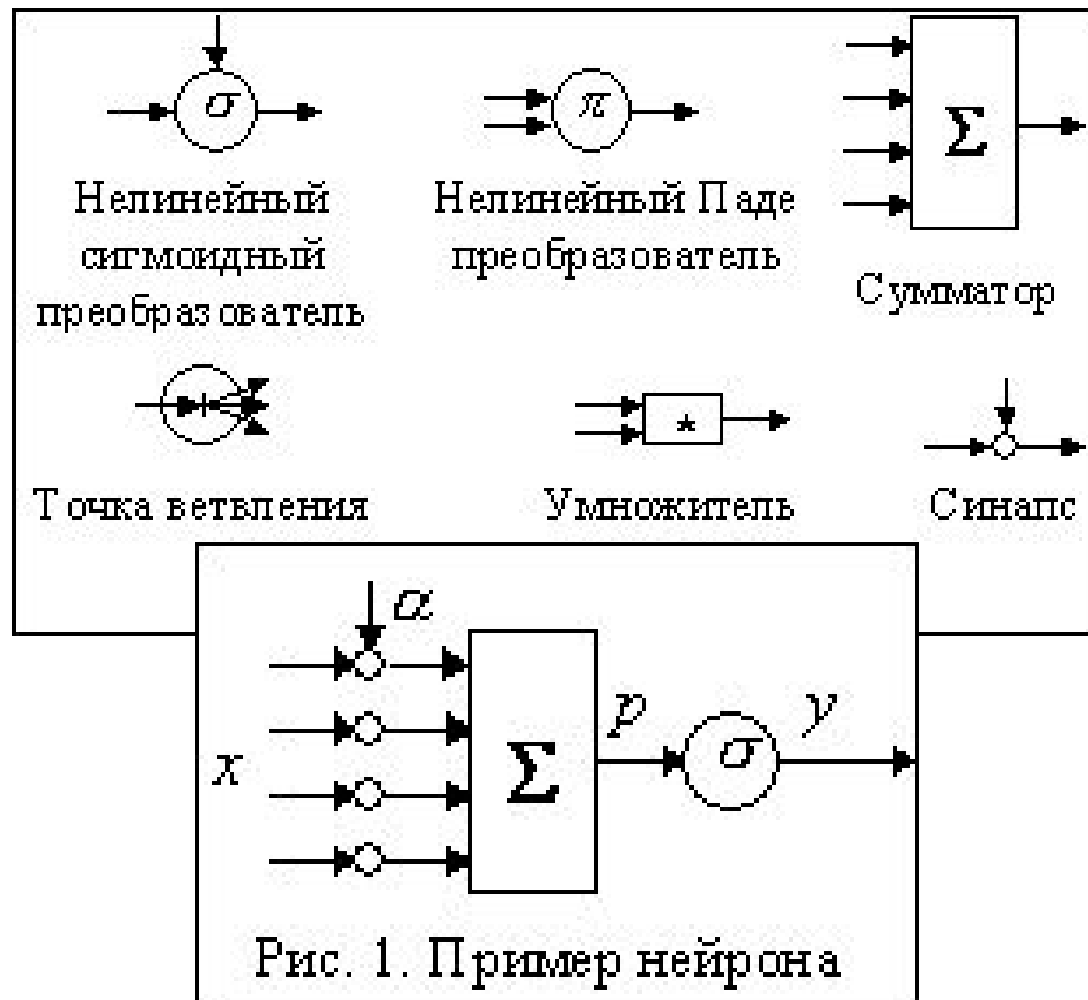
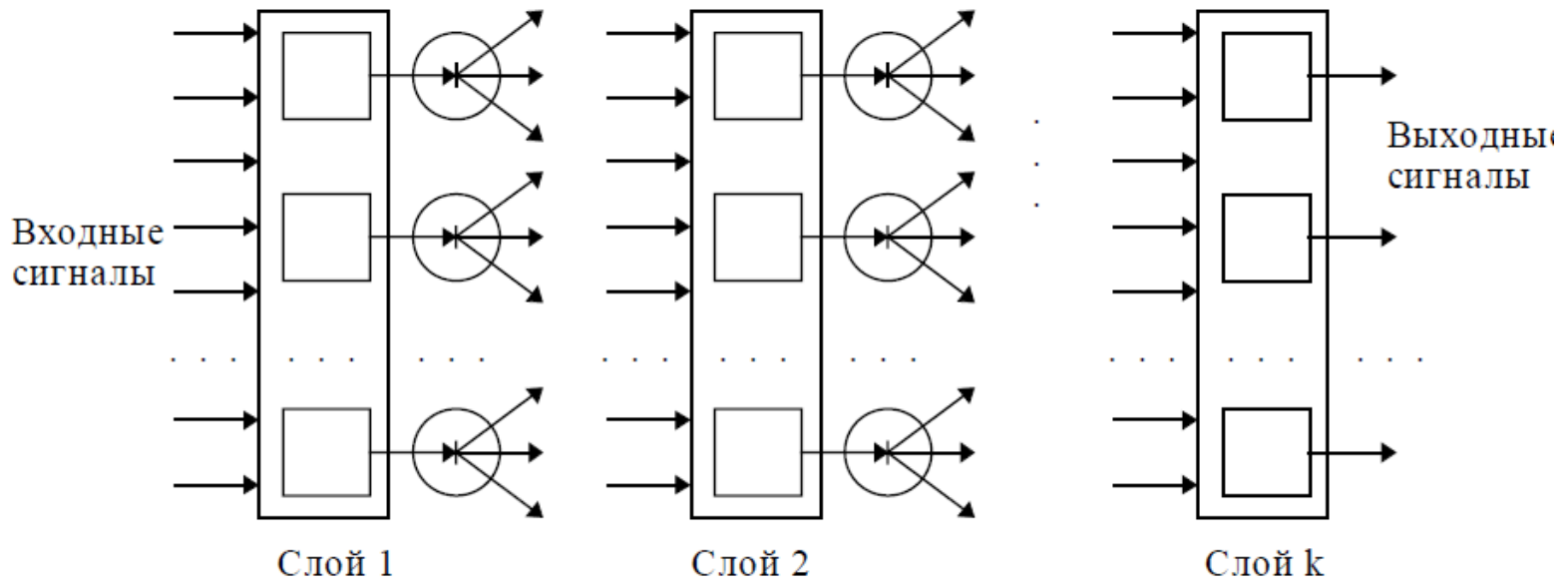
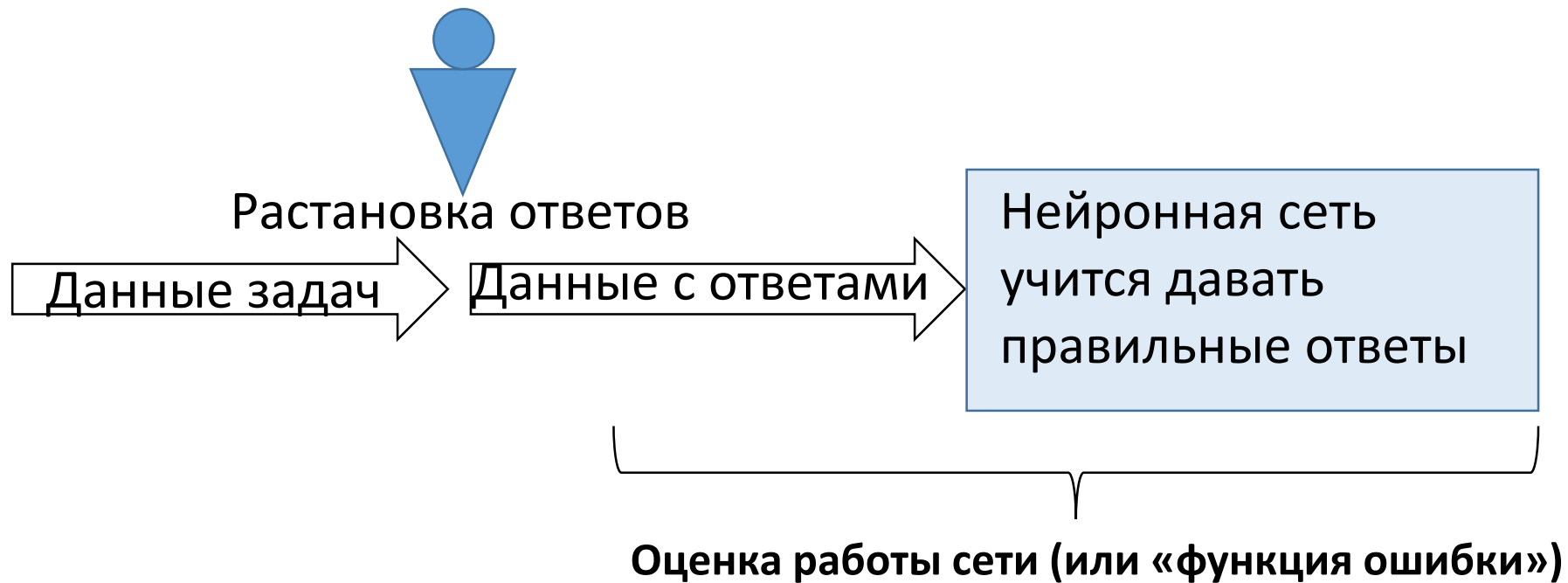


Рис. 1. Пример нейрона

Слоистая сеть

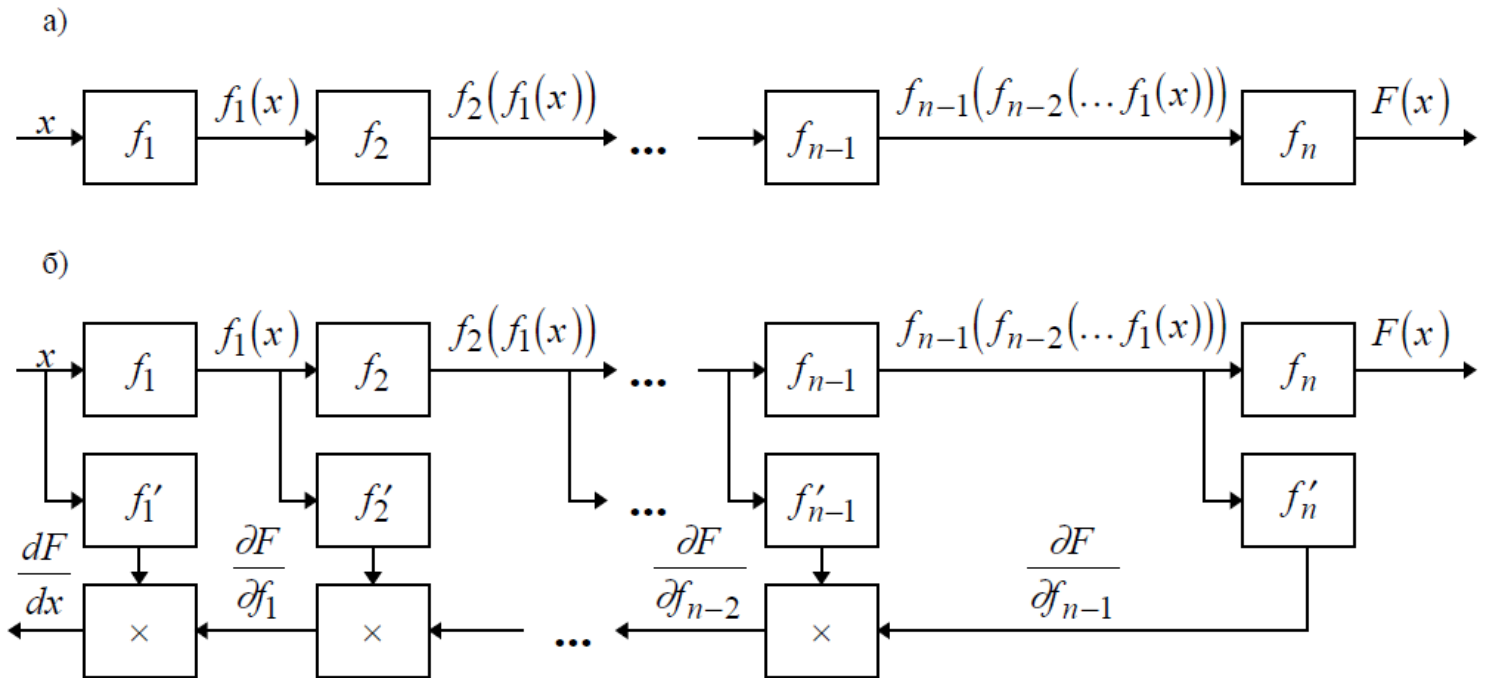


Обучение с учителем



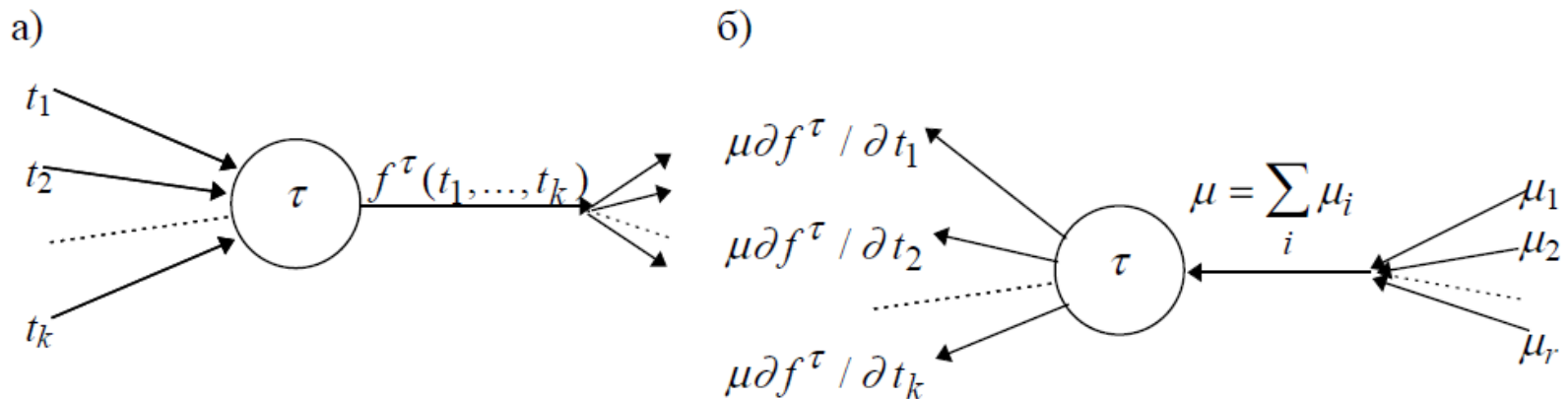
Как правило, сеть обучается путем минимизации функции ошибки (с добавлением «поощрения» за хорошую структуру сети или штрафа за избыточную сложность)

Принцип двойственности для вычисления градиента оценки (метод обратного распространения)



А для функций многих переменных добавляется ветвление при обратном функционировании

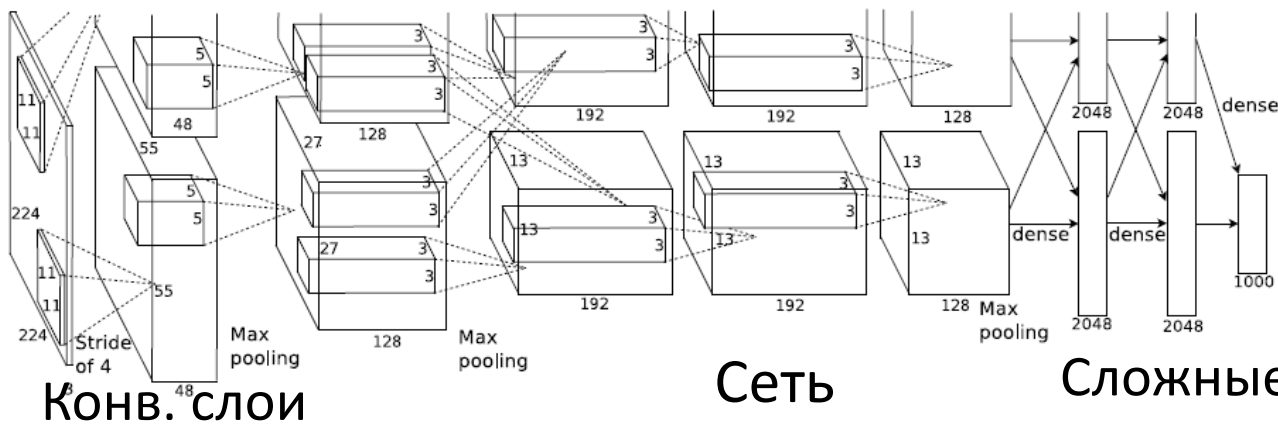
Двойственные функциональные элементы



Прохождение функционального элемента в прямом (а) и обратном (б) направлениях

Глубокое обучение конволюционных сетей для обработки изображений и сигналов

- **Глубокое** – много слоев;
- **Конволюционные сети** – обрабатывают сигналы и изображения путем их свертки с **простыми линейными фильтрами** (много иерархически выстроенных фильтров);
- Слои **простых нелинейных фильтров** («победитель получает все» на своей площадке, и т.п.).



96 обученных
фильтров

Производство знаний из данных

- **Нейросети производят знания для использования человеком, переноса в другие программные продукты и др.**
- «Именно нейросетевая технология производства знаний, по-моему, является «точкой роста», которая по-новому развернет нейроинформатику, преобразует многие разделы информатики и создаст новые.» Горбань, 1998.
- **Технология есть (давно), софт есть, но пользователей мало** – рынок требует «кнопочных решений».
- **ПОЛЬЗОВАТЕЛЬ ЛЕНИВ** – это фундаментальный принцип развития ИИ

Смена базового принципа

- Первая схема: **нейросети суммируют неявный опыт и используют его для решения задач** (можно тиражировать, как программы или устройства).
- Следующая схема: **нейросети производят знания для использования человеком, переноса в другие программные продукты и др.**
- Нарождающаяся схема: **нейросети производят знания и учат друг друга.** Они функционируют в разных условиях, накапливают разные знания и обмениваются ими.

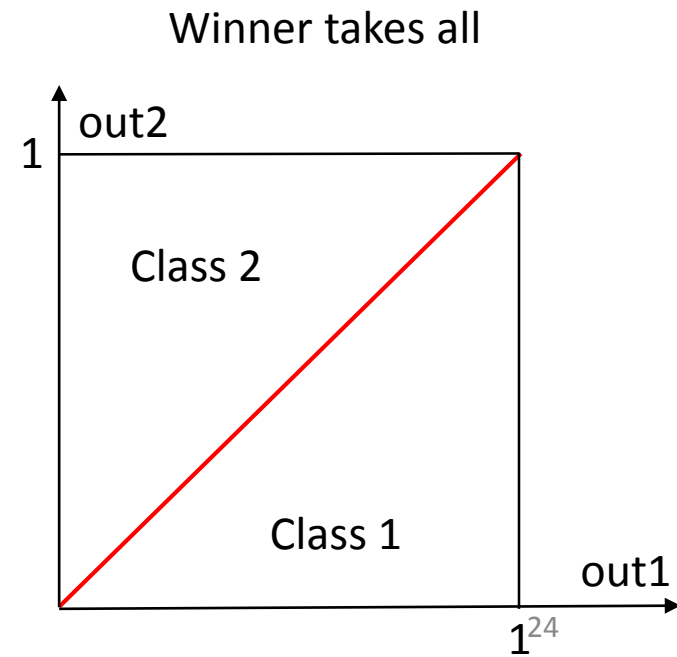
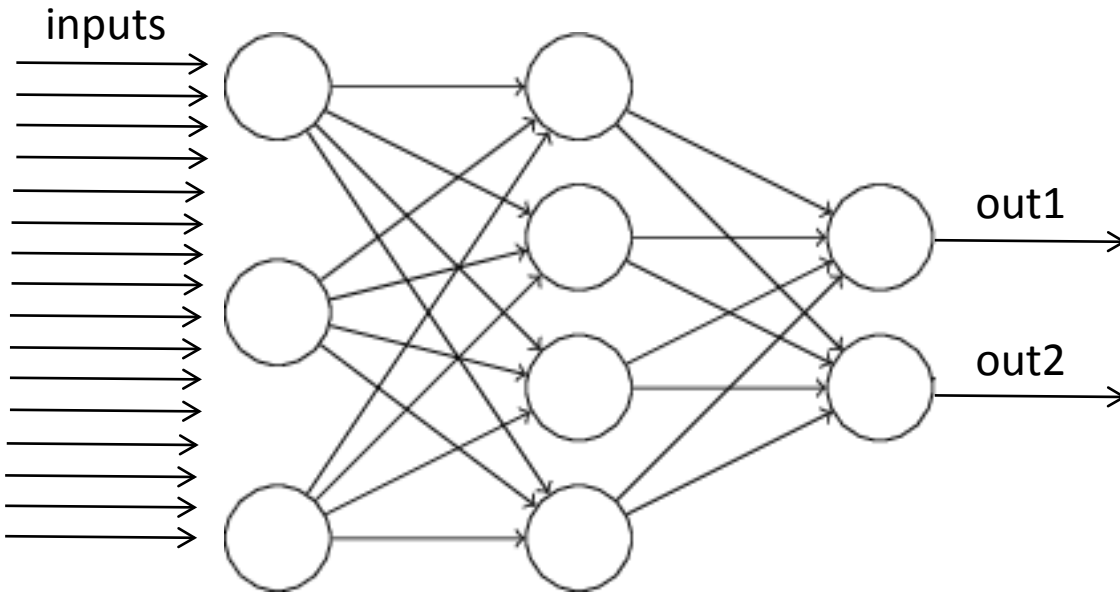
AUGMENTING ARTIFICIAL INTELLECT: A CONCEPTUAL FRAMEWORK

A paraphrase of Douglas C. Engelbart

Two papers published simultaneously in 1991

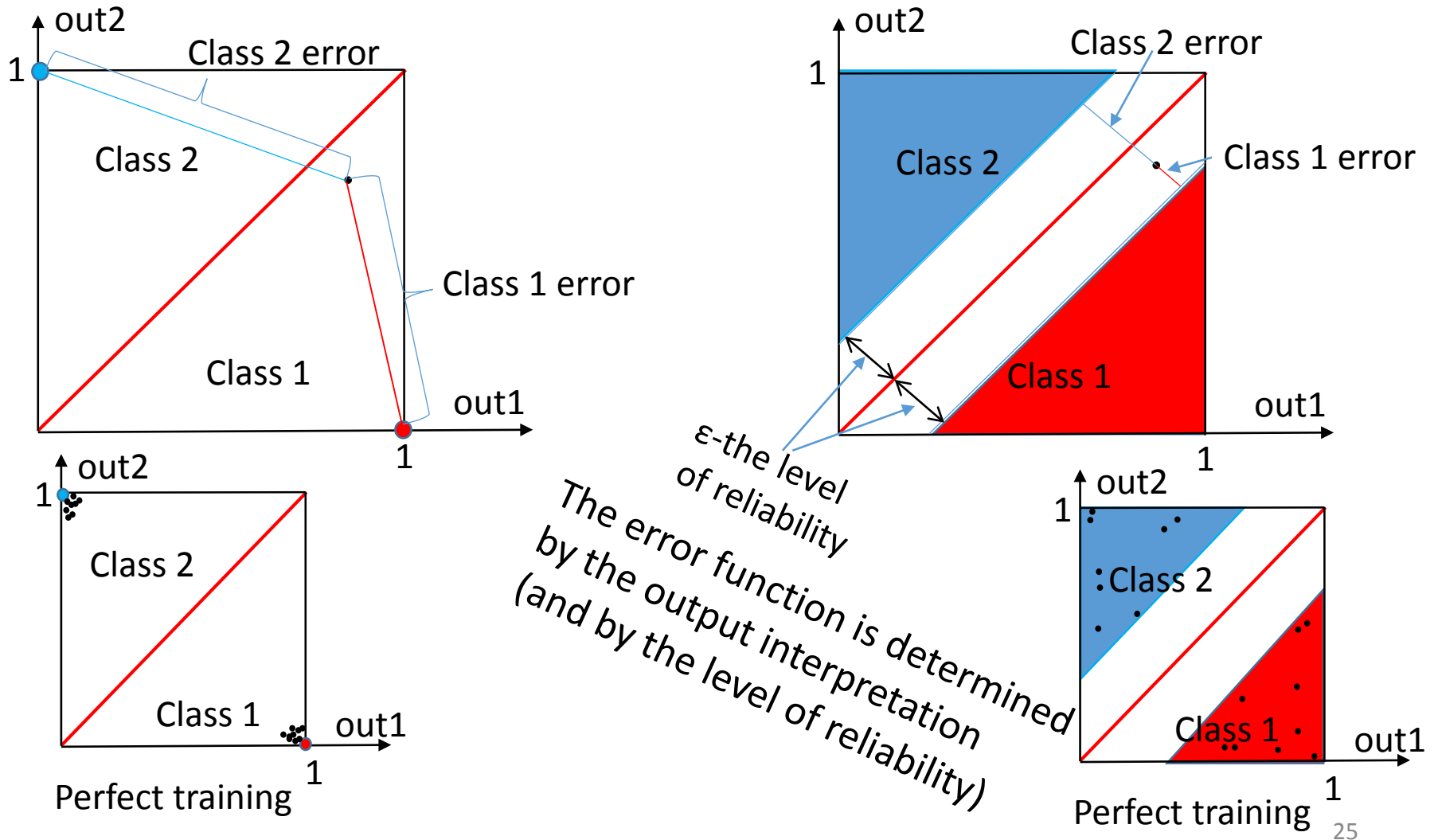
- SE Gilev, AN Gorban, EM Mirkes, Small experts and internal conflicts in learning neural networks. *Akademiia Nauk SSSR, Doklady* 320 (1), 220-223, 1991. (**GGM1991**)
- RA Jacobs, MI Jordan, SJ Nowlan, GE Hinton, Adaptive mixtures of local experts, *Neural computation* 3 (1), 79-87, 1991.

Interpretation of outputs (classifier)



Classifiers are used as a simple class of AI systems

Interpretation and error function

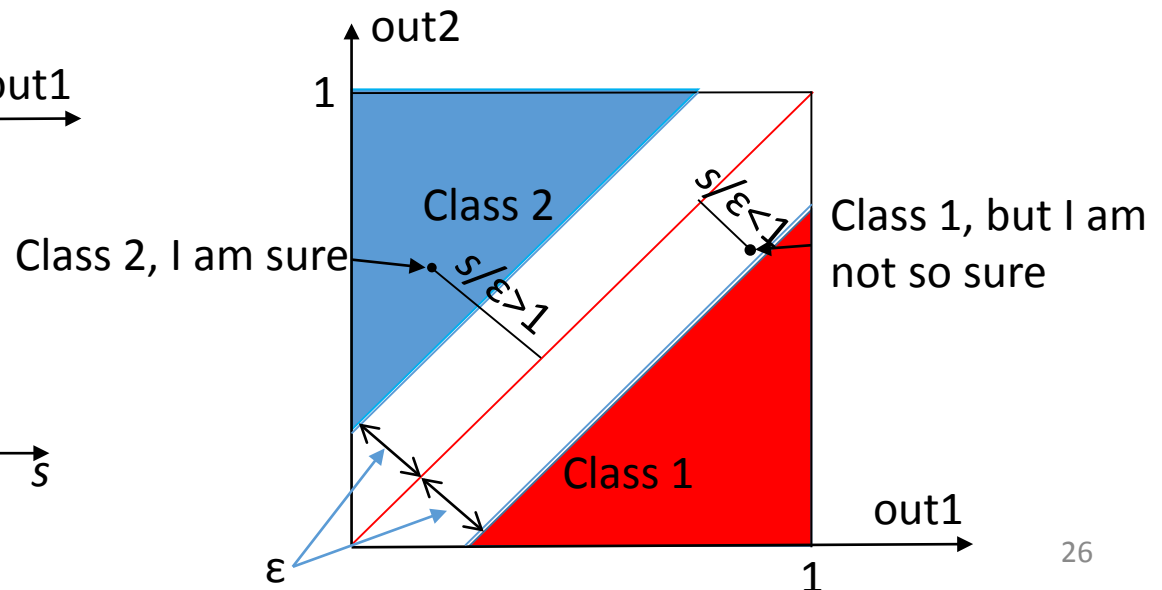
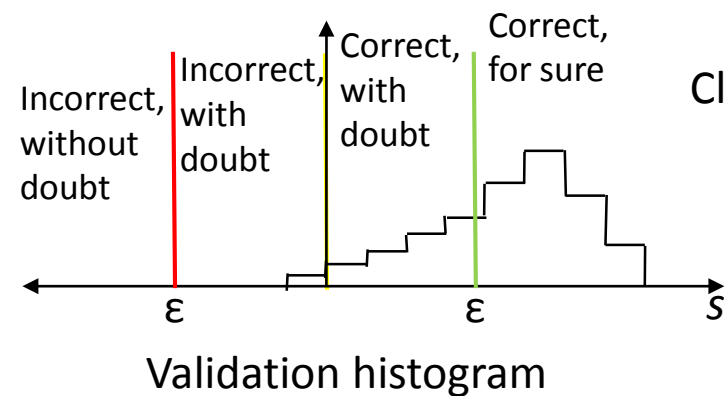
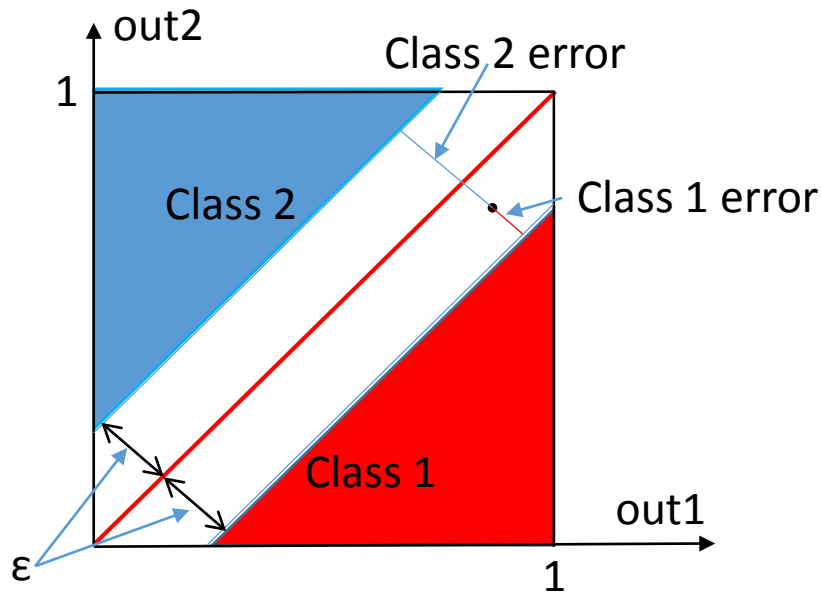


Benefits

1. Simplification of learning tasks (less neurons and less time is needed for training);

2. New useful characteristic appears: the level of self-esteem s (the distance from the diagonal):

If $s/\epsilon > 1$ then the system "is sure" in the answer (the training is assumed to be validated with given ϵ)



Idea 1

Let many agents share the problem fields.

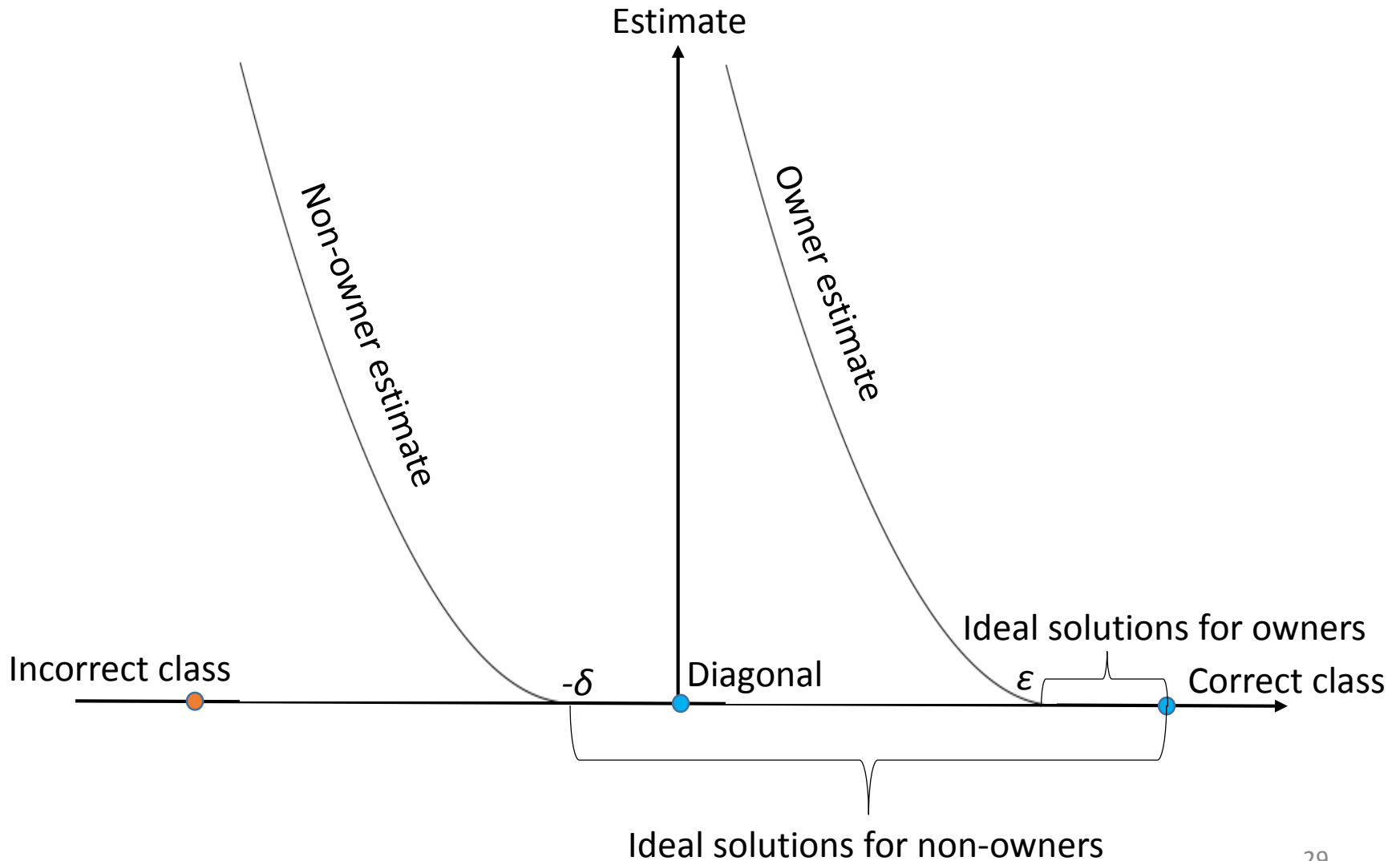
We teach agents:

- To solve **their own** problems correctly with high level of self-esteem ε ;
- To solve problems **owned by other agents** correctly or *“incorrectly but with lower level of self-esteem $\delta < \varepsilon$ ”*.

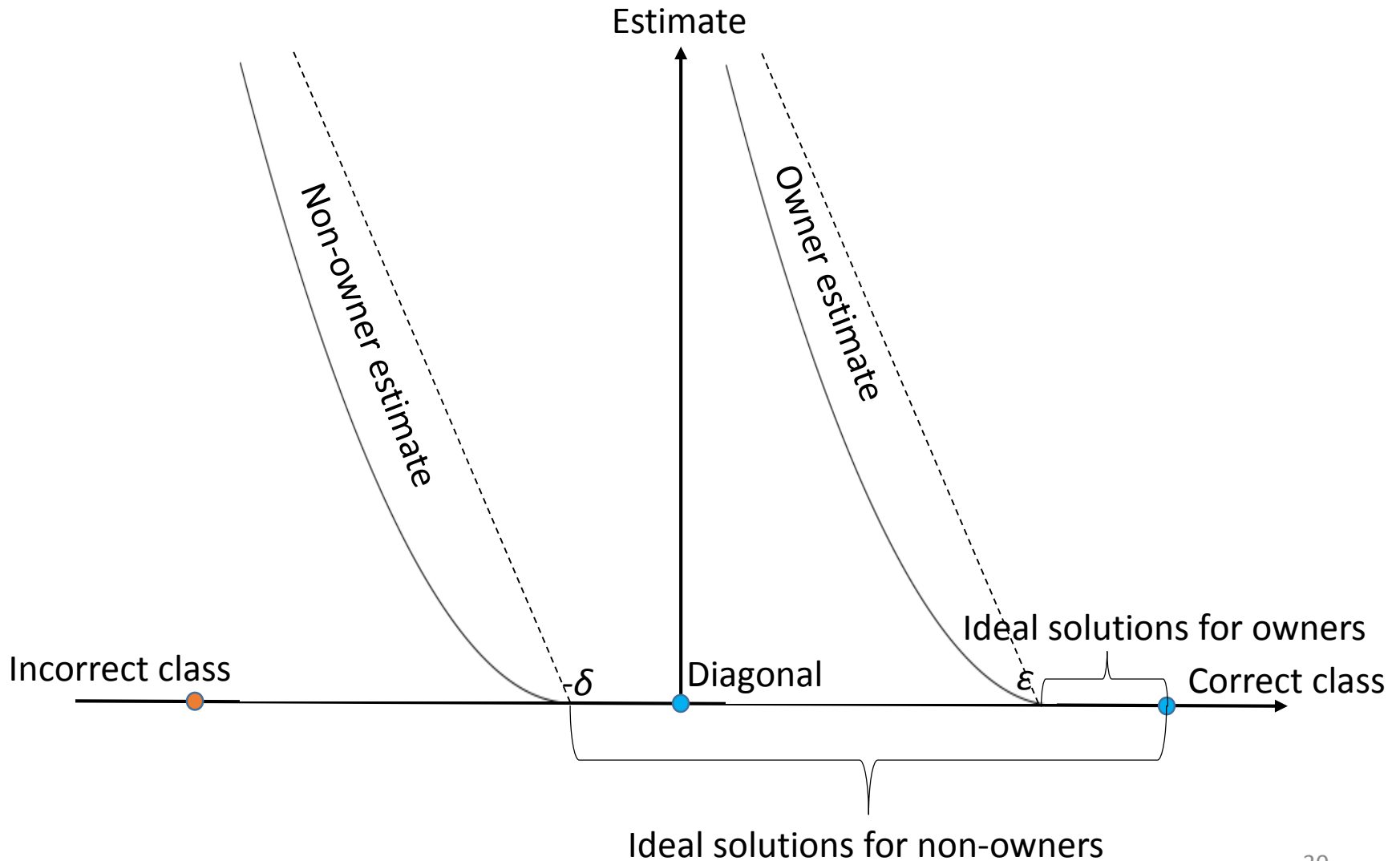
Supervising teaching

- Take a task
- Select the owner (an agent with the best correct solution, or with the smallest mistake)
- Teach the owner by a step or several steps of gradient descent of the error function with self-esteem ε ;
- Teach non-owners with a step or several steps of gradient descent of the error function with self-esteem $-\delta$ (*with $\delta < \varepsilon$*).
- Take the next task

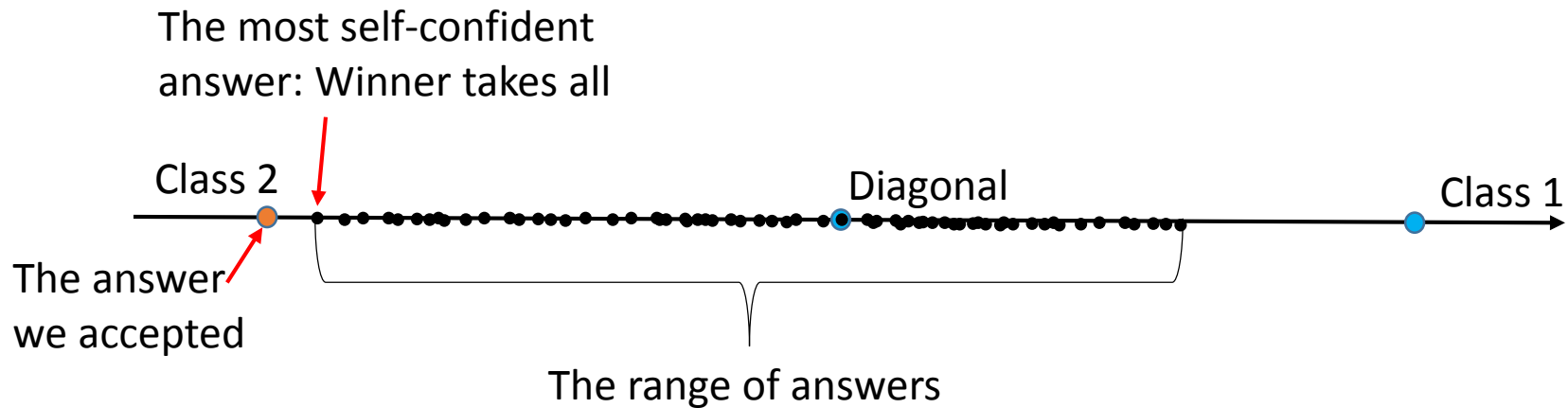
1D plots of MSE (for teaching)



A piece-wise linear estimate



Functioning: the agent with maximal self-esteem gives the final answer



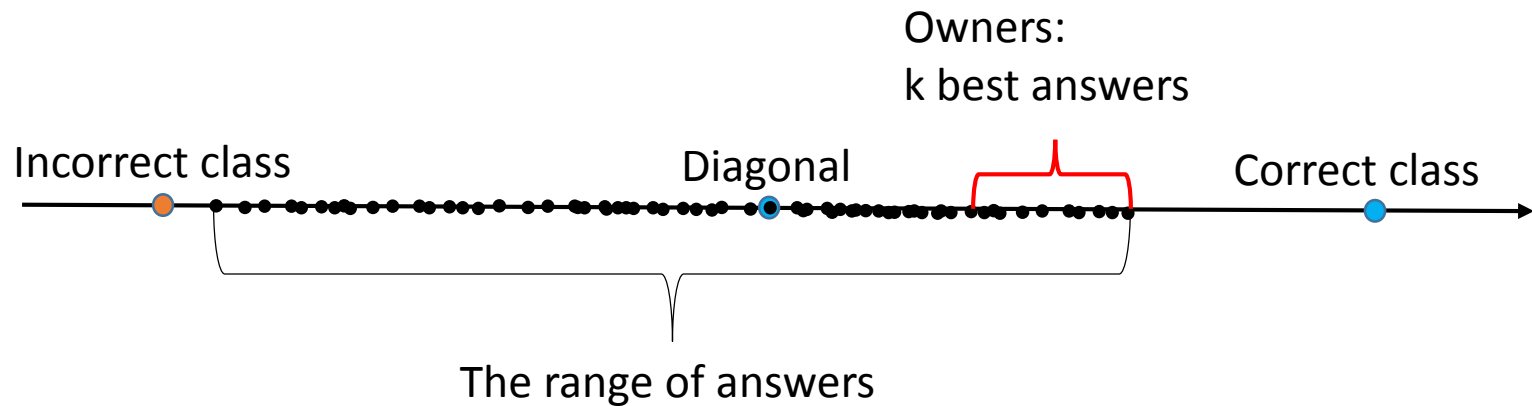
Floating classes, unsupervised dynamic coding and skills updating

In real life, the classes “float” in time and the classification model should also float.

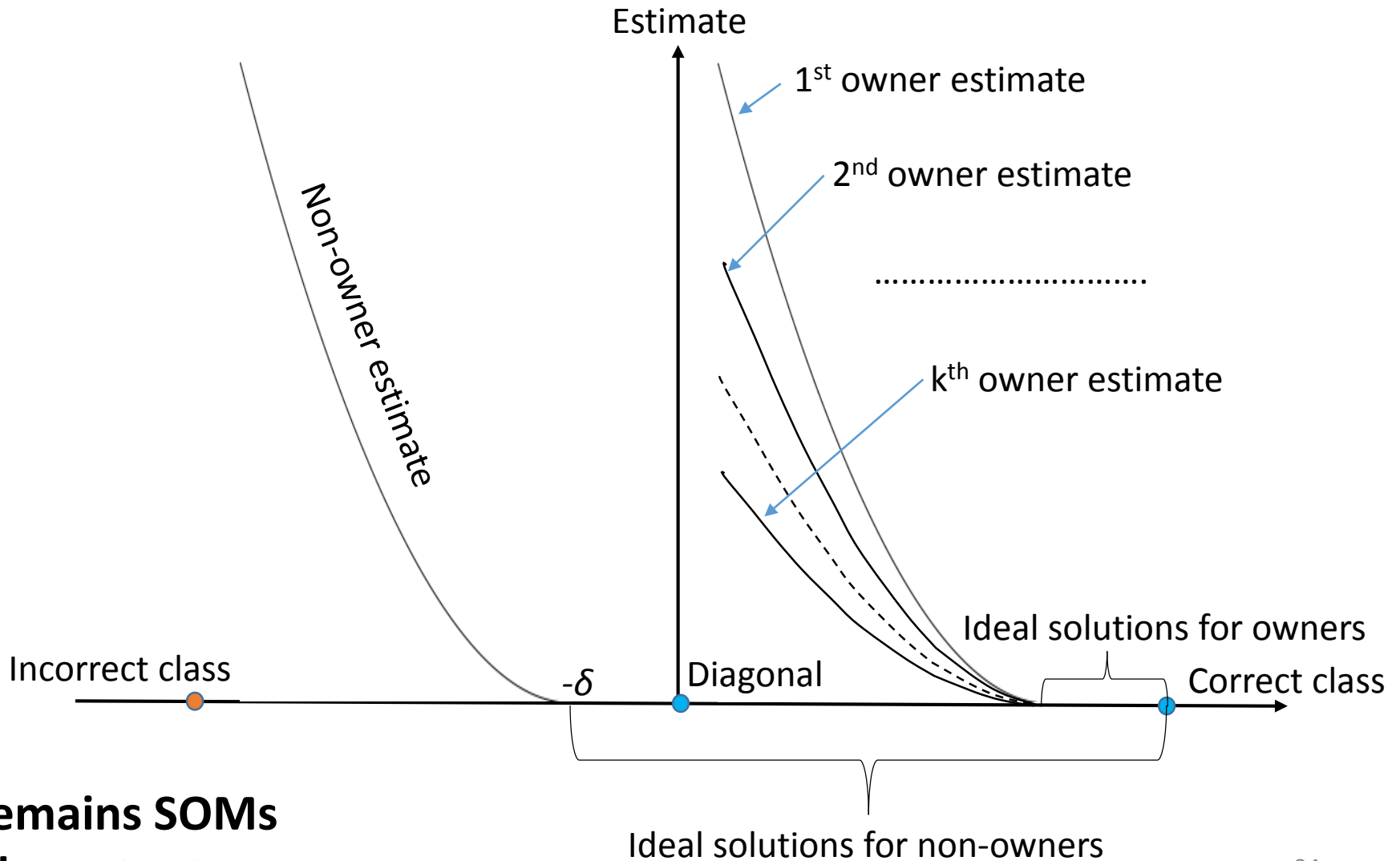
- Take a new sample;
- Label of this new example by the “winner takes all” rule;
- Use this label for a step of supervising teaching (for the owner and for non-owners).

Collective ownership

- In supervising learning the owners of a sample are k agents with k best correct solution, or with the smallest mistake.

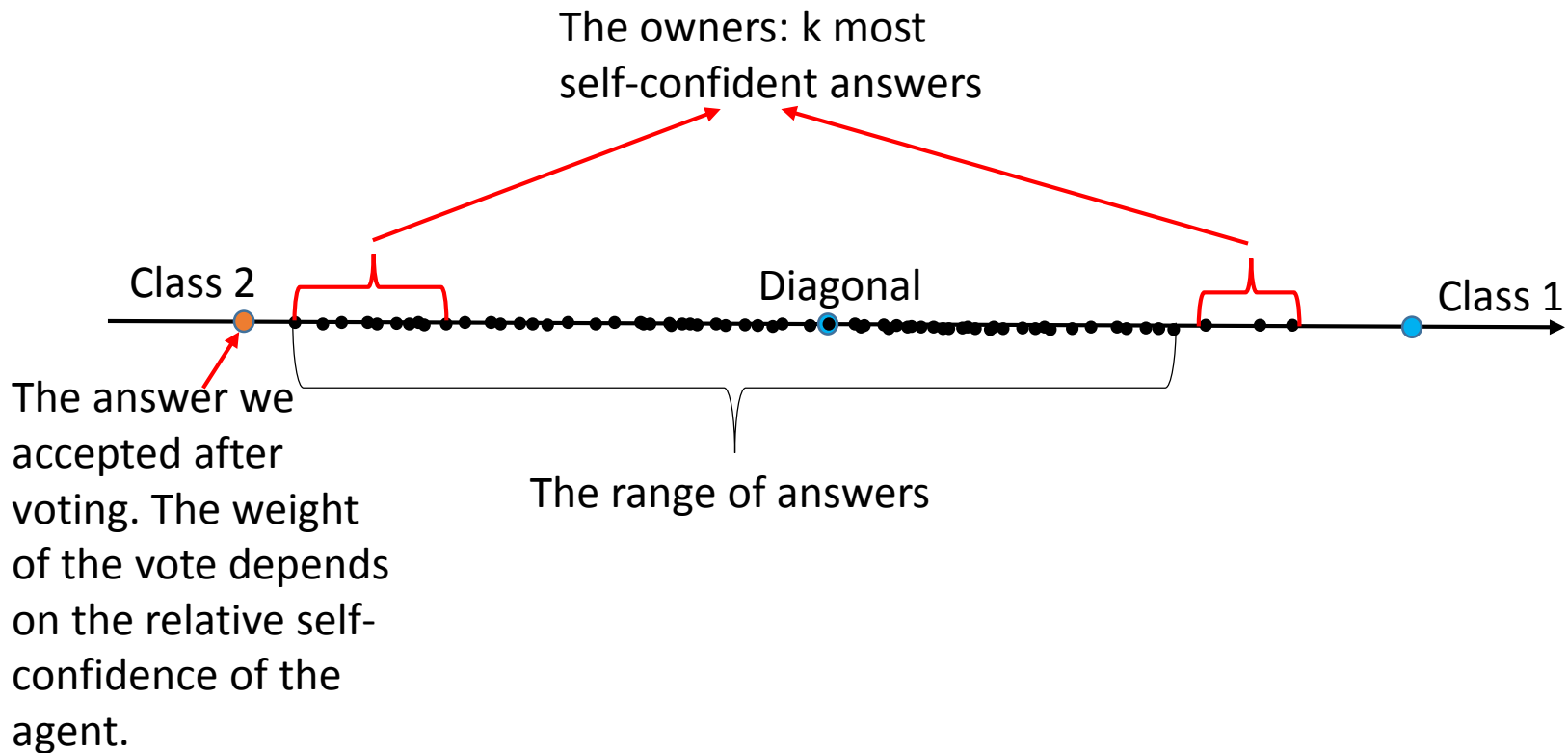


Estimates for collective ownership



**It remains SOMs
in the output space**

Functioning with collective ownership: the weighted voting of owners



Floating classes, unsupervised dynamic coding and skills updating for collective ownership

- Take a new sample;
- Find k most self-confident agents for this sample
- Label this new example by the voting of owners;
- Use this label for a step of supervising teaching (for the owners and for non-owners).

TOO MUCH ITERATIVE LEARNING

The Blessing of Dimensionality: Separation Theorems in the Thermodynamic Limit ^{*}

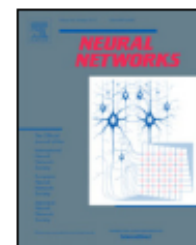
Alexander N. Gorban ^{*} Ivan Yu. Tyukin ^{**} Ilya Romanenko ^{***}

Neural Networks 94 (2017) 255–259

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet



Stochastic separation theorems

A.N. Gorban ^{a,*}, I.Y. Tyukin ^{a,b}

^a Department of Mathematics, University of Leicester, Leicester, LE1 7RH, UK

^b Department of Automation and Control Processes, Saint-Petersburg State Electrotechnical University, Saint-Petersburg, 197 376, Russia



ARTICLE INFO

Article history:

Received 1 April 2017

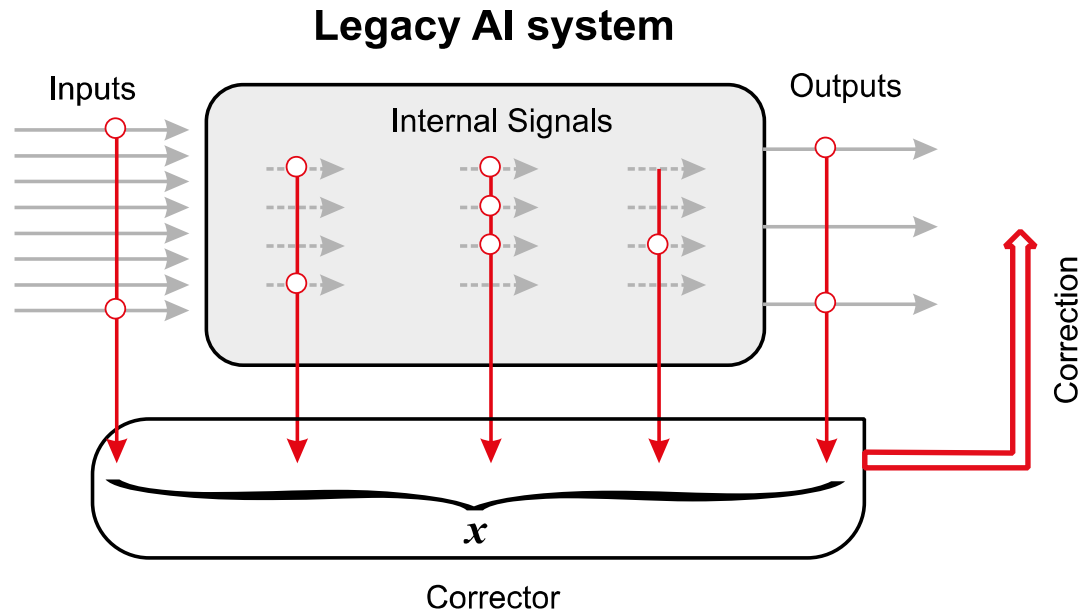
Received in revised form 16 July 2017

Accepted 21 July 2017

ABSTRACT

The problem of non-iterative one-shot and non-destructive correction of unavoidable mistakes arises in all Artificial Intelligence applications in the real world. Its solution requires robust separation of samples with errors from samples where the system works properly. We demonstrate that in (moderately) high

One-shot AI Knowledge Transfer



At the unsupervised stage the agents are considered as legacy AI systems

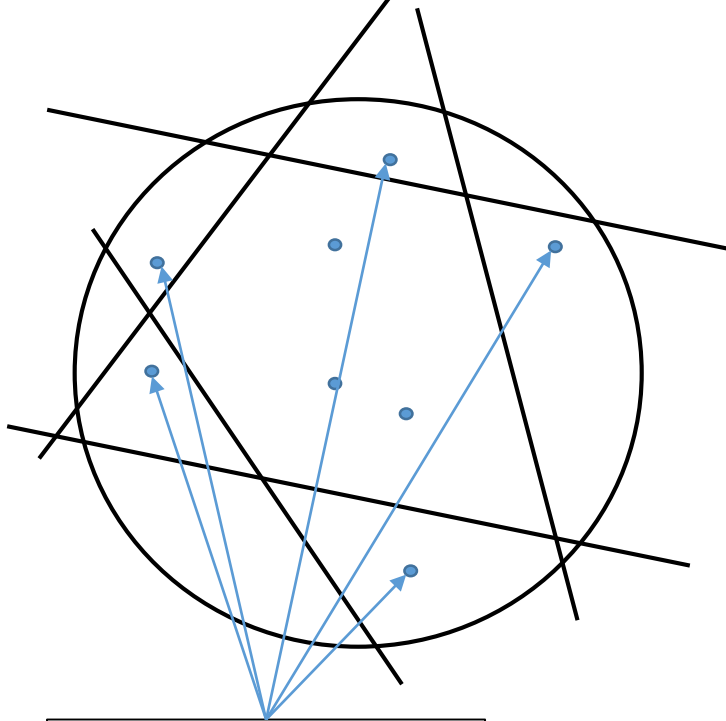
Corrector building blocks: linear functionals or small neural networks

Corrector should:

- be simple;
- not change the skills of the legacy system;
- allow fast non-iterative learning;
- allow correction of new mistakes without destroying of previous corrections.

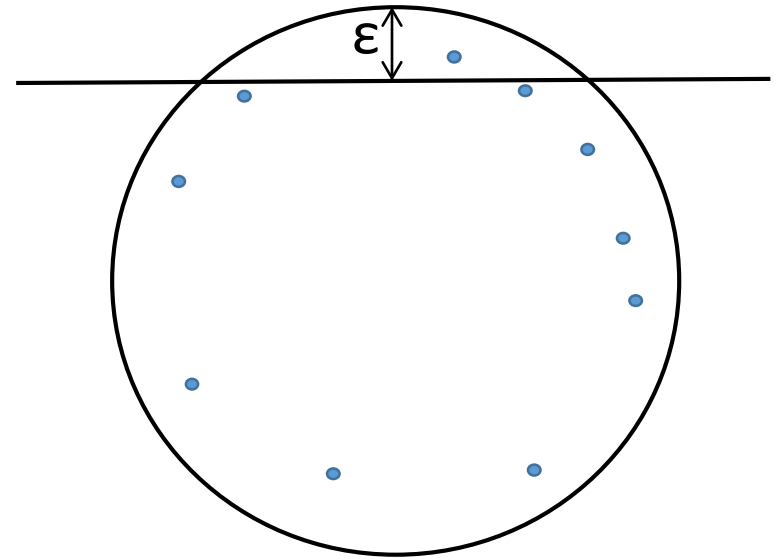
Thus, it has to separate mistakes from correctly solved examples and correct mistakes!

Stochastic separation theorems in high dimensions



Extreme points

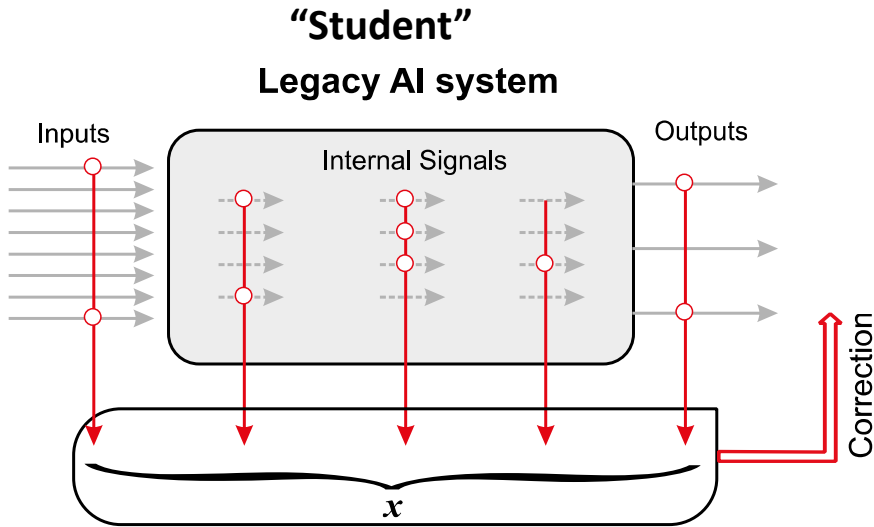
Expected ε IS NOT small in high dimension?



In high dimension, with high probability in an exponentially large random set all points are extreme ones???

AI Knowledge Transfer from teacher to a student

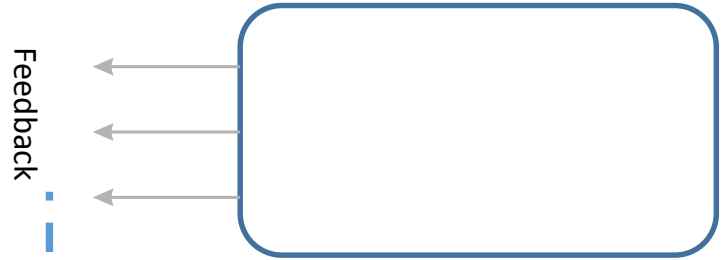
A student *Legacy AI system* processes some input signals, produces internal representations $x \in R^n$ of the input and returns *some* outputs



Elementary "Additional" and Efficient computationally decisions on x

A teacher *Legacy AI system* processes the input signals related to the input seen by the student and returns expert outputs

"Teacher"
Legacy AI system



The input Data "seen" by the Student generates the set

$$\mathcal{M} = \{x_1, x_2, \dots, x_M\}, x_i \in R^n$$

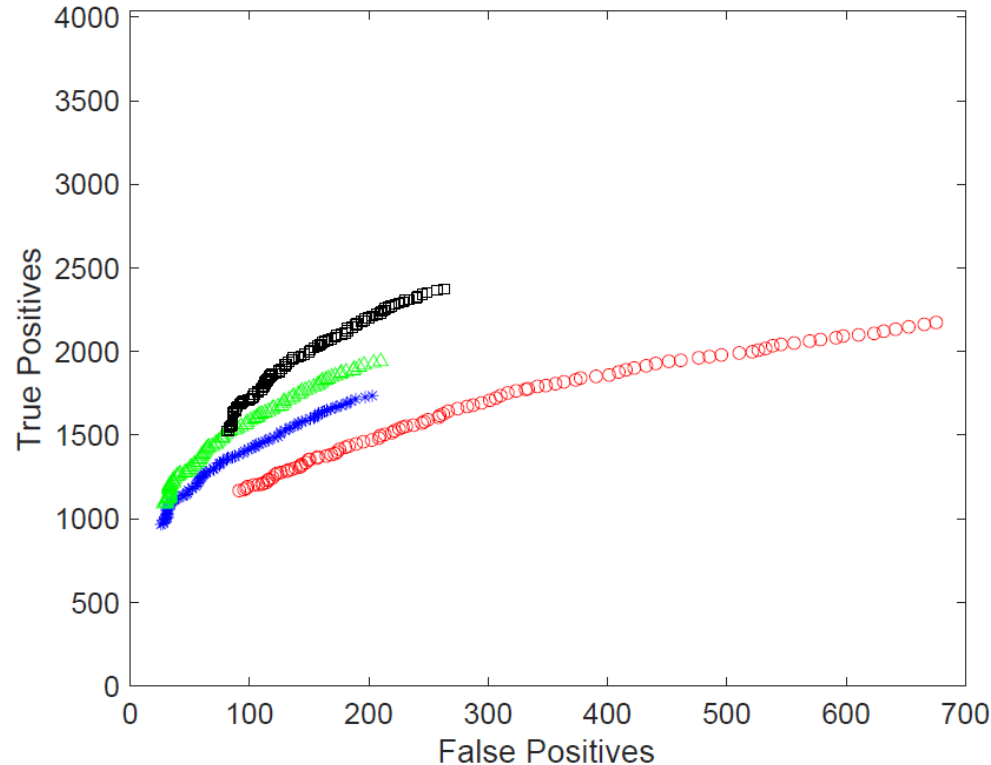
of internal representations

Knowledge transfer between AIs

In our experiments:

- The teacher AI, AI_t , was modelled by a deep Convolutional Network, ResNet 18 with circa 11M trainable parameters.
- The teacher network was trained on a “teacher” dataset comprised of 5.2M non-pedestrian (negatives), and 600K pedestrian (positives) images.
- The student AI, AI_s , was modelled by a linear classifier with 2016 trainable parameters.
- The values of these parameters were the result of AIs training on
- a student” dataset, a sub-sample of the “teacher” dataset comprising of 130K negatives and 55K positives.
- This choice of AI_s and AI_t systems enabled us to emulate interaction between edge-based AIs and their more powerful counterparts that could be deployed on larger servers or computational clouds.

Knowledge transfer between AIs: Student's errors before and after teaching



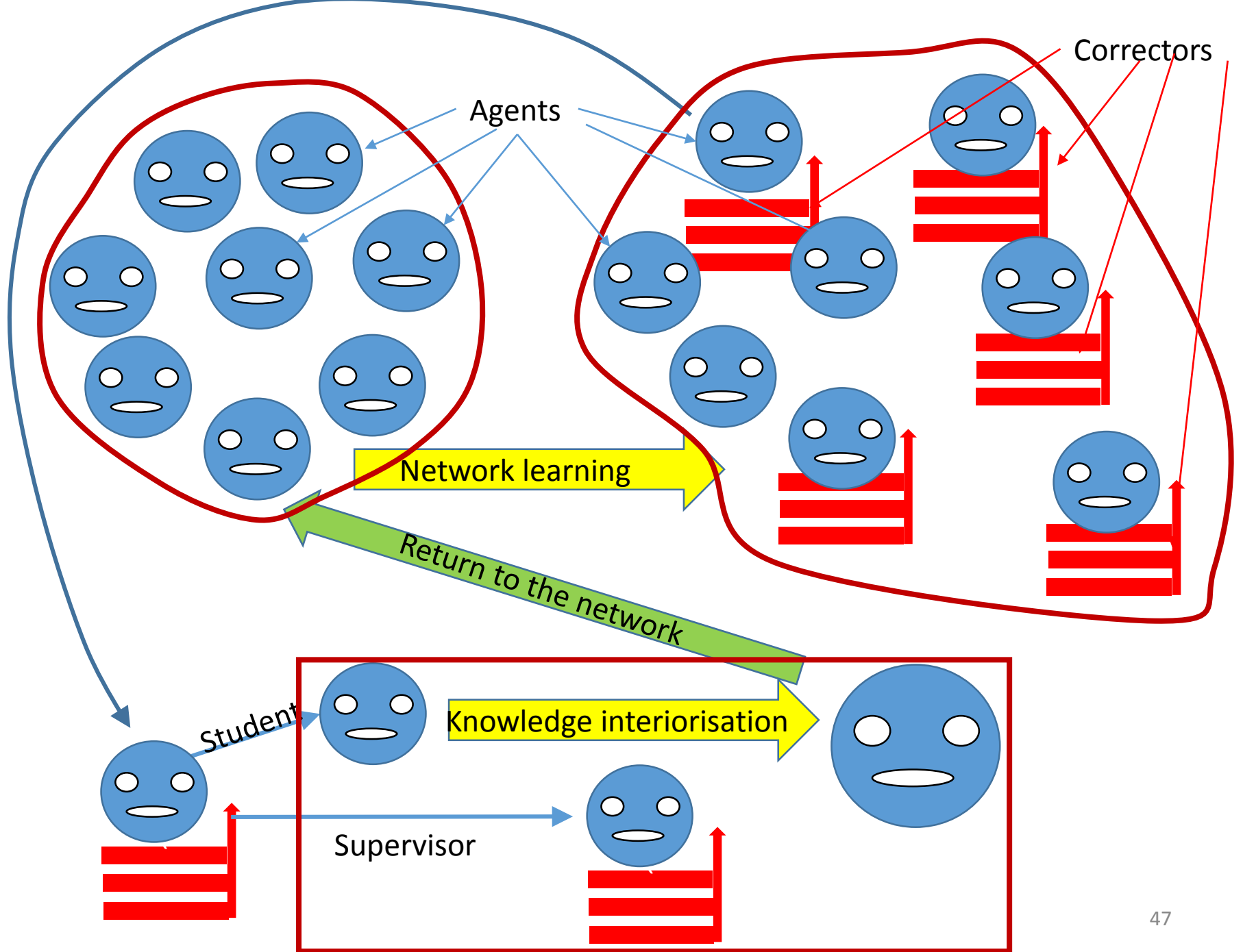
- Red circles show true positives as a function of false positives for the original student linear classifier.
- Blue stars and Green triangles correspond to AIs after correction of false positives by the teacher (two different algorithms or errors clustering).
- Black squares correspond to AIs after correction of false positive followed by correction of false negative.

Non-iterative knowledge transfer from the owners conference to all agents, who find the sample interesting

- Take a new sample;
- Find q ($q > k$) most self-confident agents for this sample (the agents who find this sample “interesting”; k most self-confident of them are owners)
- Label this new example by the voting of owners;
- Create a non-iterative corrector for the agents who are interested in this sample but make a mistake.
- If a “human supervisor” makes labelling then it can be used for corrections of all the agents who are interested in this sample but make a mistake.

Interiorisation of knowledge

- After collection of correctors, the interiorisation of the collected knowledge can be used.
- Interiorisation of knowledge is its incorporation into agent's inner structure.
- Interiorisation can be organised as an iterative supervised learning procedure that uses the agent with correctors for sample labelling as the supervisor.
- Interiorisation is performed for all agents separately and independently, in parallel.
- After the individual interiorisation procedure, the agents return to the network learning.



Results and problems

- +AI with self-esteem are efficient even for standard learning tasks;
- +The multiagent systems with self-esteem can solve the problems of growing complexity;
- +These systems combine the supervising pre-learning and self-learning of the networks with corrections and knowledge transfer;
- +Interiorisation of knowledge is its incorporation into agent's inner structure for further learning and work.
- ?The scalability of the method to big and very big heterogeneous problem spaces has to be analysed.

Some references

- Хайкин С. Нейронные сети: полный курс, 2-е издание. Издательский дом Вильямс; 2008.
- Schmidhuber J. Deep learning in neural networks: An overview. Neural networks. 2015 Jan 31;61:85-117.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems 2012 (pp. 1097-1105).

Some References (2)

- Горбань АН, Россиев ДА, Нейронные сети на персональном компьютере. Новосибирск: наука; 1996.
- Gorban A.N., Tyukin I.Y., Stochastic Separation Theorems, Neural Networks, 94, October 2017, 255-259.
- A.N. Gorban, I.Y. Tyukin, I. Romanenko, The Blessing of Dimensionality: Separation Theorems in the Thermodynamic Limit, IFAC-PapersOnLine 49-24 (2016), 064–069.
- Knowledge Transfer Between Artificial Intelligence Systems, arXiv:1709.01547 [cs.AI]
- One-Trial Correction of Legacy AI Systems and Stochastic Separation Theorems, arXiv:1610.00494 [stat.ML]