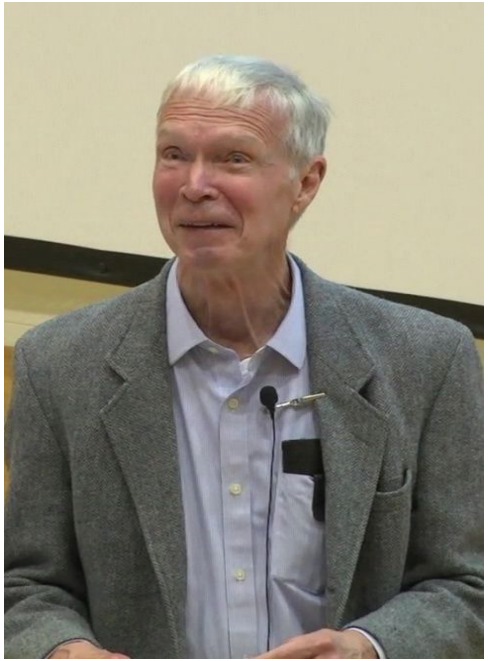
УНИВЕРСИТЕТ ЛОБАЧЕВСКОГО

Почему Нобелевскую премию по физике 2024 г. дали за исследования в области машинного обучения?

Н.Ю. Золотых
ННГУ им. Н.И. Лобачевского
Директор института ИТММ
Директор исследовательского центра в сфере ИИ

XXI Школа «Нелинейные волны»
Нижний Новгород, 5–11 ноября 2024 г.

Нобелевские лауреаты по физике 2024



Джон Хопфилд
род. 15 июля 1933

«За фундаментальные открытия и изобретения, обеспечивающие машинное обучение с помощью искусственных нейронных сетей»

Джеффри Хинтон
род. 6 декабря 1947

John Hopfield

ПОДПИСАТЬСЯ

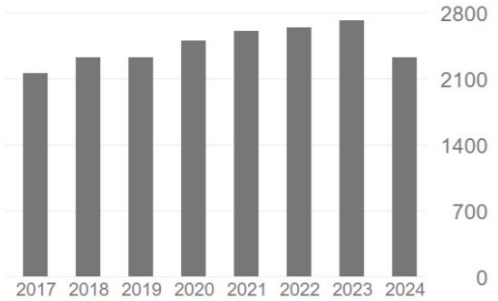
Professor, Princeton University
Подтвержден адрес электронной почты в домене princeton.edu - Главная страница

Neural Networks AI Neuroscience Systems Biology Semiconductor Physics

НАЗВАНИЕ	ПРОЦИТИРОВАНО	ГОД
Neural networks and physical systems with emergent collective computational abilities. JJ Hopfield Proceedings of the national academy of sciences 79 (8), 2554-2558	28021	1982
Neurons with graded response have collective computational properties like those of two-state neurons. JJ Hopfield Proceedings of the national academy of sciences 81 (10), 3088-3092	9618	1984
“Neural” computation of decisions in optimization problems JJ Hopfield, DW Tank Biological cybernetics 52 (3), 141-152	8994	1985
From molecular to modular cell biology LH Hartwell, JJ Hopfield, S Leibler, AW Murray Nature 402 (Suppl 6761), C47-C52	4623	1999

Процитировано ПРОСМОТРЕТЬ ВСЕ

	Все	Начиная с 2019 г.
Статистика цитирования	91551	15190
h-индекс	94	33
i10-индекс	176	75



Общий доступ ПРОСМОТРЕТЬ ВСЕ

0 статей 2 статьи

недоступно доступно

Google Академия

Geoffrey Hinton

ПОДПИСАТЬСЯ

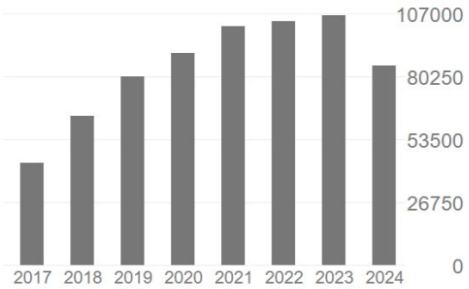
Emeritus Prof. Computer Science, University of Toronto
Подтвержден адрес электронной почты в домене cs.toronto.edu - Главная страница

machine learning psychology artificial intelligence cognitive science computer science

НАЗВАНИЕ	ПРОЦИТИРОВАНО	ГОД
Imagenet classification with deep convolutional neural networks A Krizhevsky, I Sutskever, GE Hinton Advances in neural information processing systems 25	165162 *	2012
Deep learning Y LeCun, Y Bengio, G Hinton Nature 521 (7553), 436-44	86433	2015
Learning internal representations by error-propagation DE Rumelhart, GE Hinton, RJ Williams Parallel Distributed Processing: Explorations in the Microstructure of …	55181 *	1986
Dropout: a simple way to prevent neural networks from overfitting N Srivastava, G Hinton, A Krizhevsky, I Sutskever, R Salakhutdinov The journal of machine learning research 15 (1), 1929-1958	53506	2014
Visualizing data using t-SNE L van der Maaten, G Hinton Journal of Machine Learning Research 9 (Nov), 2579-2605	48711	2008

Процитировано ПРОСМОТРЕТЬ ВСЕ

	Все	Начиная с 2019 г.
Статистика цитирования	864759	570357
h-индекс	187	137
i10-индекс	484	372



Общий доступ ПРОСМОТРЕТЬ ВСЕ

1 статья 8 статей

недоступно доступно

На основе финансирования

Премия Тьюринга по информатике 2018

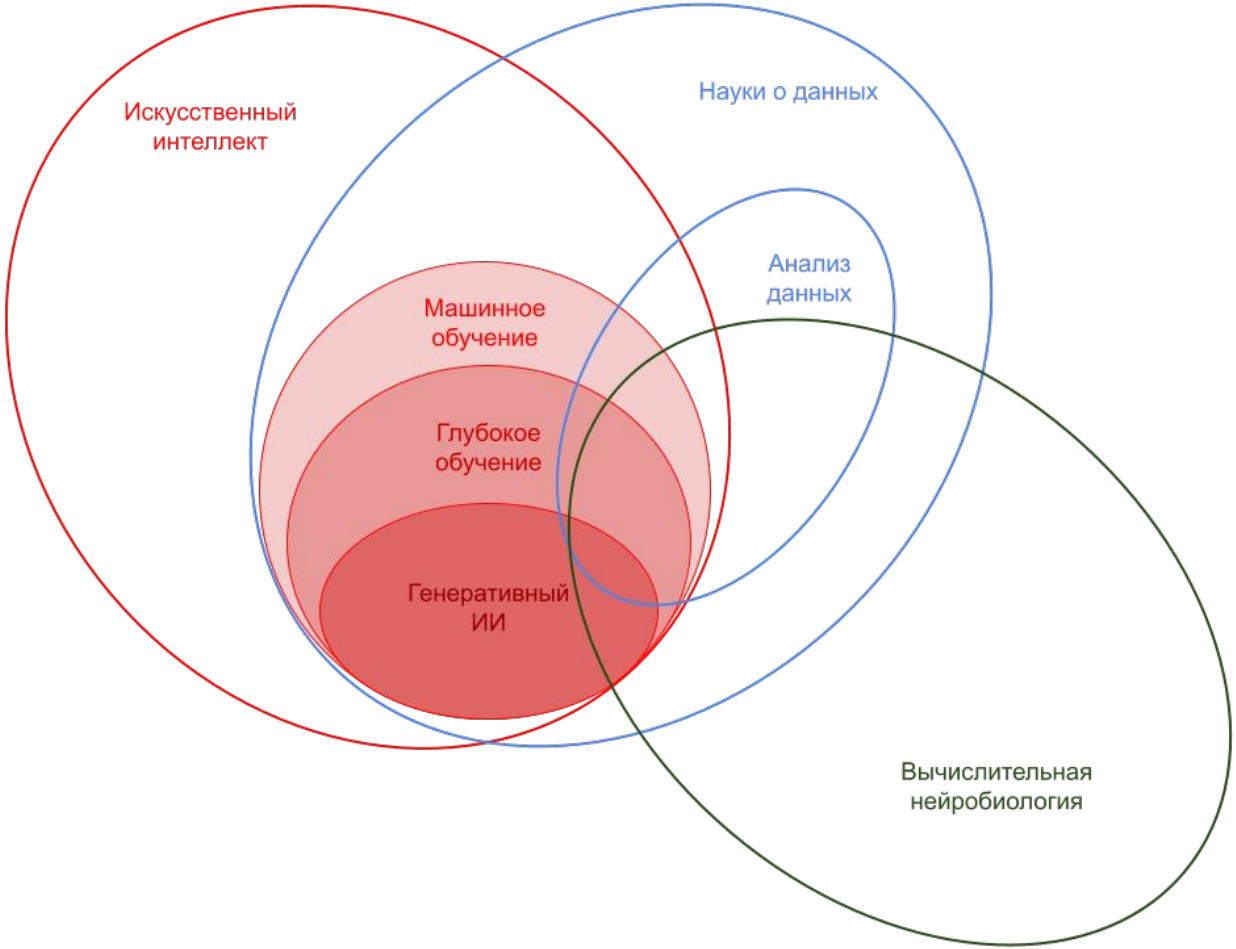
«За концептуальные и инженерные прорывы, сделавшие глубокие нейросети краеугольным компонентом в вычислительной технике»

Йошуа Бенжио, род. 1964

Джеффри Хинтон, род. 1947

Ян Лекун, род. 1960

1. История и основные концепции машинного обучения


Искусственный интеллект
Науки о данных
Анализ данных
Машинное обучение
Глубокое обучение
Генеративный ИИ
Вычислительная нейробиология

Некоторые задачи машинного обучения

- **Обучение с учителем:**
  - классификация
  - регрессия
  - предсказание временного ряда
  - …
- **Обучение без учителя:**
  - кластеризация
  - визуализация
  - понижение размерности
  - ...
- **Обучение с подкреплением**
- **…**

Обучение с учителем

$$f^*: X \to Y \quad \text{или} \quad \mathrm{P}(y \mid x)$$

Функция f^* известна только на обучающей выборке:

$$(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(N)}, y^{(N)})$$

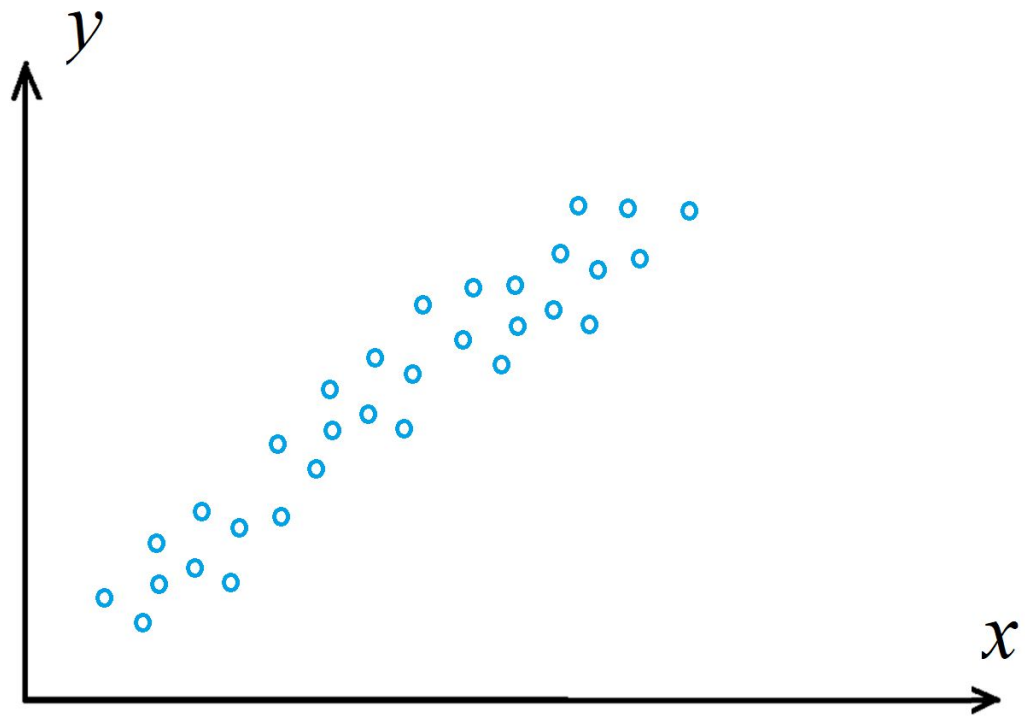
Надо найти $f(x) \approx f^*(x)$, в частности:

$$f(x^{(1)}) \approx y^{(1)}, f(x^{(2)}) \approx y^{(2)}, \ldots, f(x^{(N)}) \approx y^{(N)}$$

т.е. обучение с учителем – это *задача аппроксимации*.

- Классификация: Y – конечное (не очень большое)
- Восстановление регрессии: $Y = \boldsymbol{R}$
- Предсказания временного ряда: $Y = \{\text{временные ряды}\}$

Как y связан с x?



Как решать задачу обучения с учителем?

Стандартный подход – *метод минимизации эмпирического риска*

Штраф $L(y, y^*) = L(f(x), y^*)$

Например,

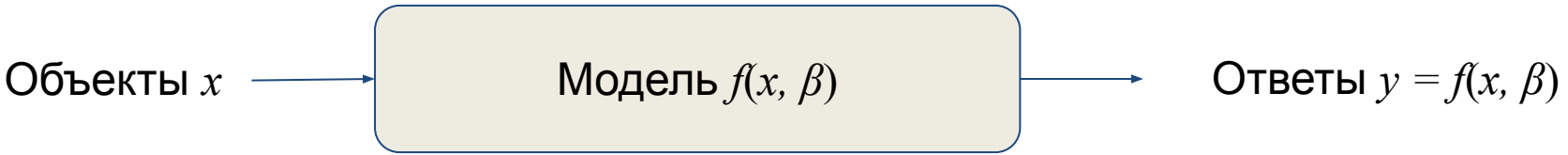
- квадратичный штраф: $L(y, y^*) = (y - y^*)^2$
- индикатор ошибки: $L(y, y^*) = I(y \neq y^*)$
- кросс-энтропия (logloss): $L(p, y^*) = -\log p_{y^*}$

Ищем параметры β неизвестной функции $f(x, \beta)$:

$$R(\beta) = \sum_{i=1\ldots N} L(f(x^{(i)}, \beta), y^{(i)}) \longrightarrow \min$$

$x^{(1)}, \ldots, x^{(N)}$

$y^{(1)}, \ldots, y^{(N)}$

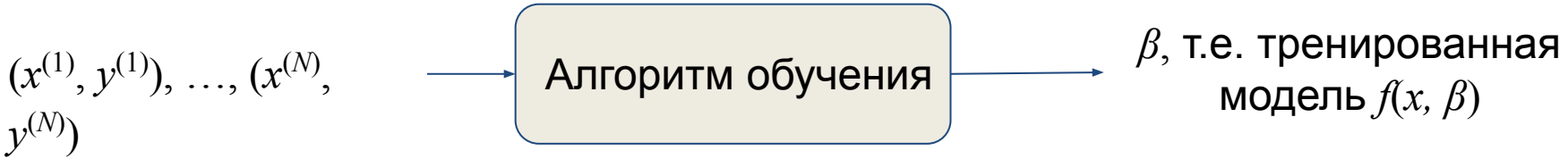


Обучение с учителем

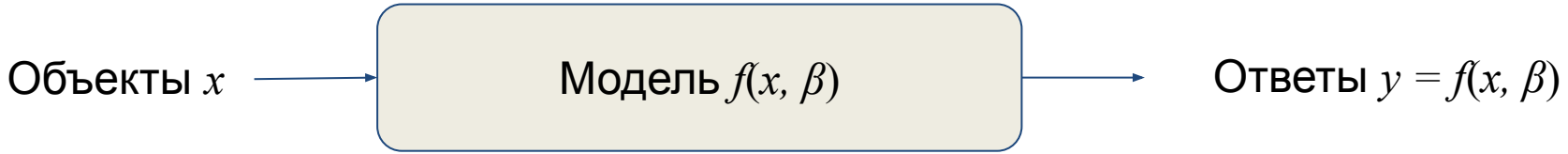
Имеется обучающая выборка: $(x^{(1)}, y^{(1)}); (x^{(2)}, y^{(2)}); \ldots; (x^{(N)}, y^{(N)})$

Надо найти $f(x) = f(x, \beta)$, так, чтобы $f(x^{(1)}) \approx y^{(1)}, f(x^{(2)}) \approx y^{(2)}, \ldots, f(x^{(N)}) \approx y^{(N)}$

Fit (настройка, обучение, learning, training)



Predict (предсказание, inference)



Три вехи и три составных части ML

$x \rightarrow y$

Вектор → Скаляр	Линейная регрессия, логистическая регрессия, MLP, SVM, RF, GBT, …
Структура → Скаляр	CNN, AlexNet, ResNet, word2vec, ...
Структура → Структура	seq2seq, LSTM, GAN, VAE, BERT, GPT, ...

Основные вехи развития нейронных сетей

Начало

- W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, 1943
- F. Rosenblatt, Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms, 1962
- M. Minsky, S. Papert, Perceptrons: an introduction to computational geometry, 1969

Метод обратного распространения ошибки

- S. Linnainmaa, The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors 1970
- А.И. Галушкин, Синтез многослойных систем распознавания образов, 1974
- P. J. Werbos, Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Science, 1974
- D.E. Rumelhart, **G.E. Hinton**, R.J. Williams, Learning Internal Representations by Error Propagation, 1985

Глубокие сети

- А.Г. Ивахненко, В.Г. Лапа, Кибернетические предсказывающие устройства, 1965
- **G.E. Hinton,** Learning multiple layers of representation, 2007

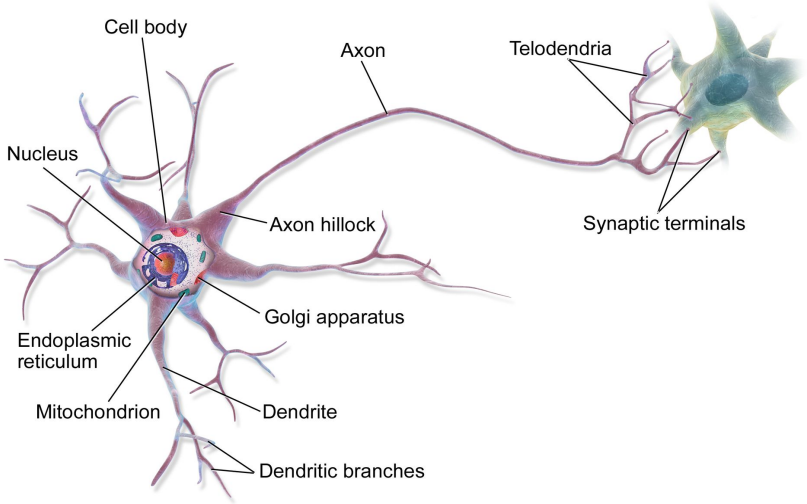
Сверточные сети

- K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, 1980
- Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, 1998
- A. Krizhevsky, I. Sutskever, **G.E. Hinton,** Imagenet classification with deep convolutional neural networks, 2012

Генеративные сети

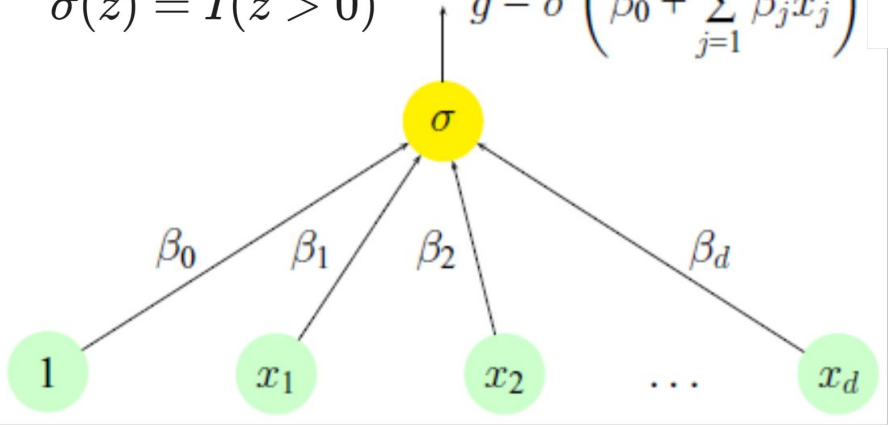
- **G.E. Hinton**, S. Osindero, Y.W. Teh, A fast learning algorithm for deep belief nets, 2006
- D.P. Kingma, M. Welling, Auto-Encoding Variational Bayes 2013
- I. Goodfellow et. al., Generative adversarial networks, 2014
- A. Vaswani et. al., Attention is all you need, 2017

Нейрон и модель МакКаллока-Питтса

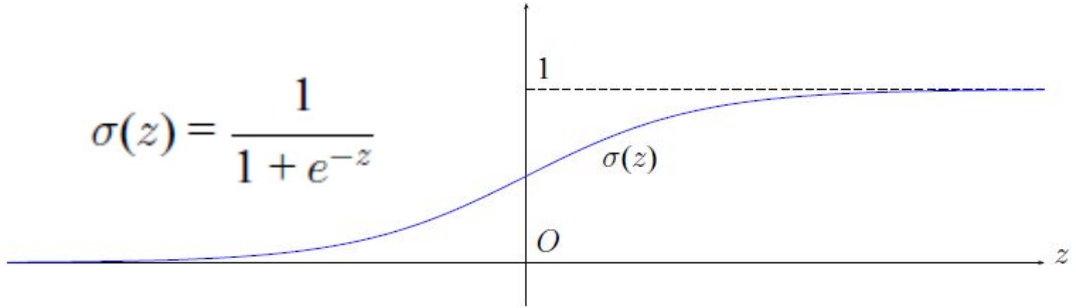


$$\sigma(z) = I(z > 0)$$

$$g = \sigma\left(\beta_0 + \sum_{j=1}^{d} \beta_j x_j\right)$$

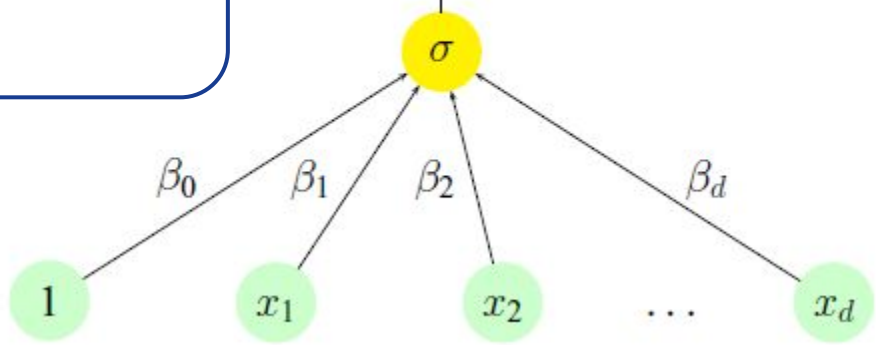


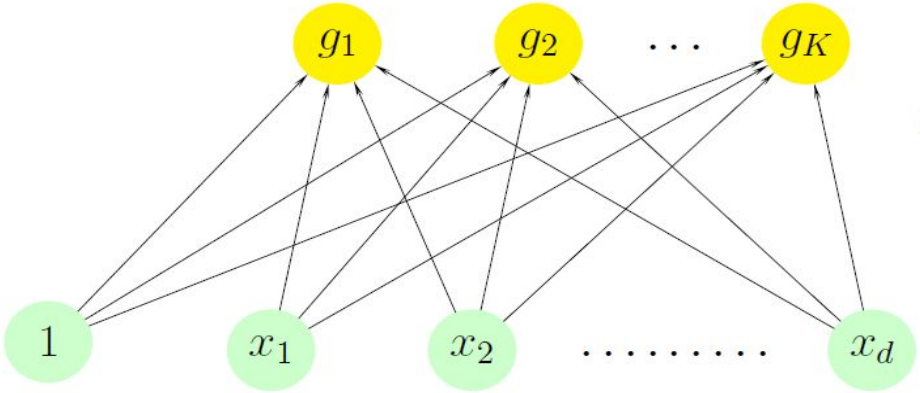
$$\sigma(z) = \frac{1}{1 + e^{-z}}$$



Логистическая функция и softmax

$$g = \sigma\left(\beta_0 + \sum_{j=1}^{d} \beta_j x_j\right) \approx \Pr(Y = 1 \,|\, X = x)$$

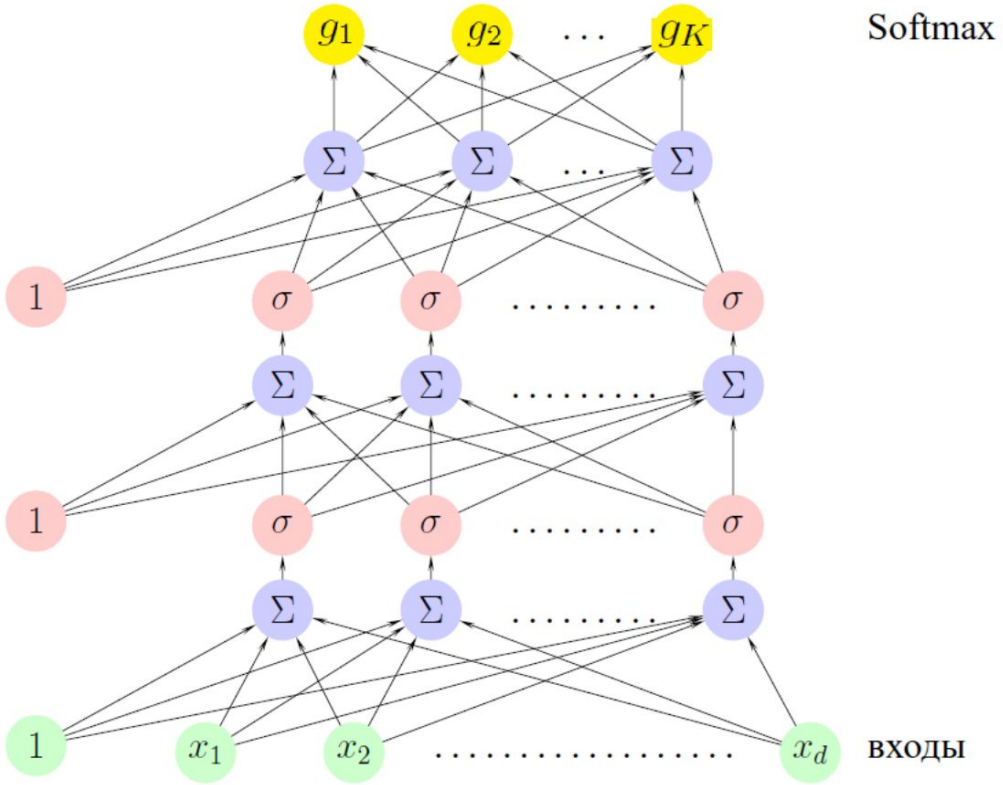




$$g_k = \frac{\exp\left(\beta_{k0} + \sum_{j=1}^{d} \beta_{kj} x_j\right)}{\sum_{\ell=1}^{K} \exp\left(\beta_{\ell 0} + \sum_{j=1}^{d} \beta_{\ell j} x_j\right)} \approx \Pr(k \,|\, x)$$

$$(k = 1, 2, \ldots, K)$$

Полносвязная нейронная сеть



BackProp: Обучение нейронной сети

- Штраф – сумма квадратов для задачи восстановления регрессии:

$$R(w) = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\frac{1}{2}\left(y^{(i)} - f(x^{(i)})\right)^2}_{R^{(i)}} \to \min$$

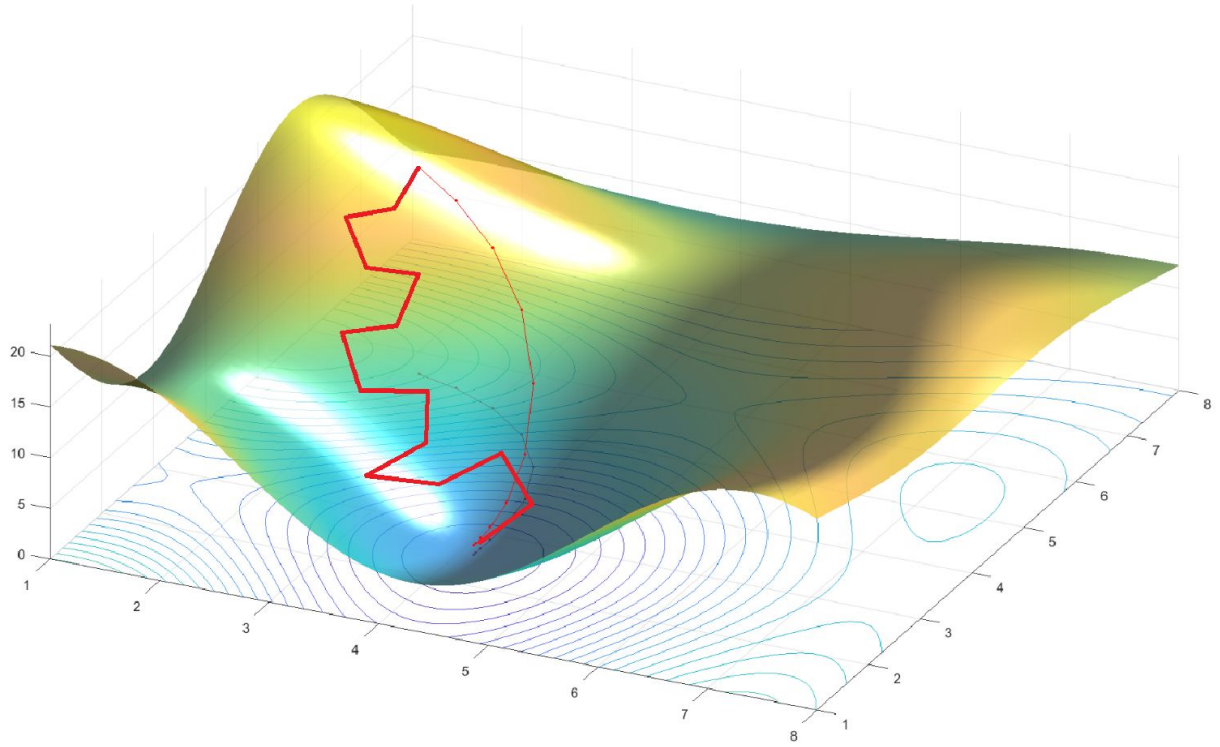
- Штраф – кросс-энтропия (logloss) для задачи классификации:

$$R(w) = -\frac{1}{N}\sum_{i=1}^{N} \underbrace{\sum_{k=1}^{K} I(y^{(i)} = k) \ln g_k(x^{(i)})}_{R^{(i)}} \to \min$$

Для решения задачи минимизации используем *алгоритм стохастического градиентного спуска*

BackPropagation – это алгоритм вычисления компонент градиен $\partial R^{(i)}/\partial w$

Стохастический градиентный спуск



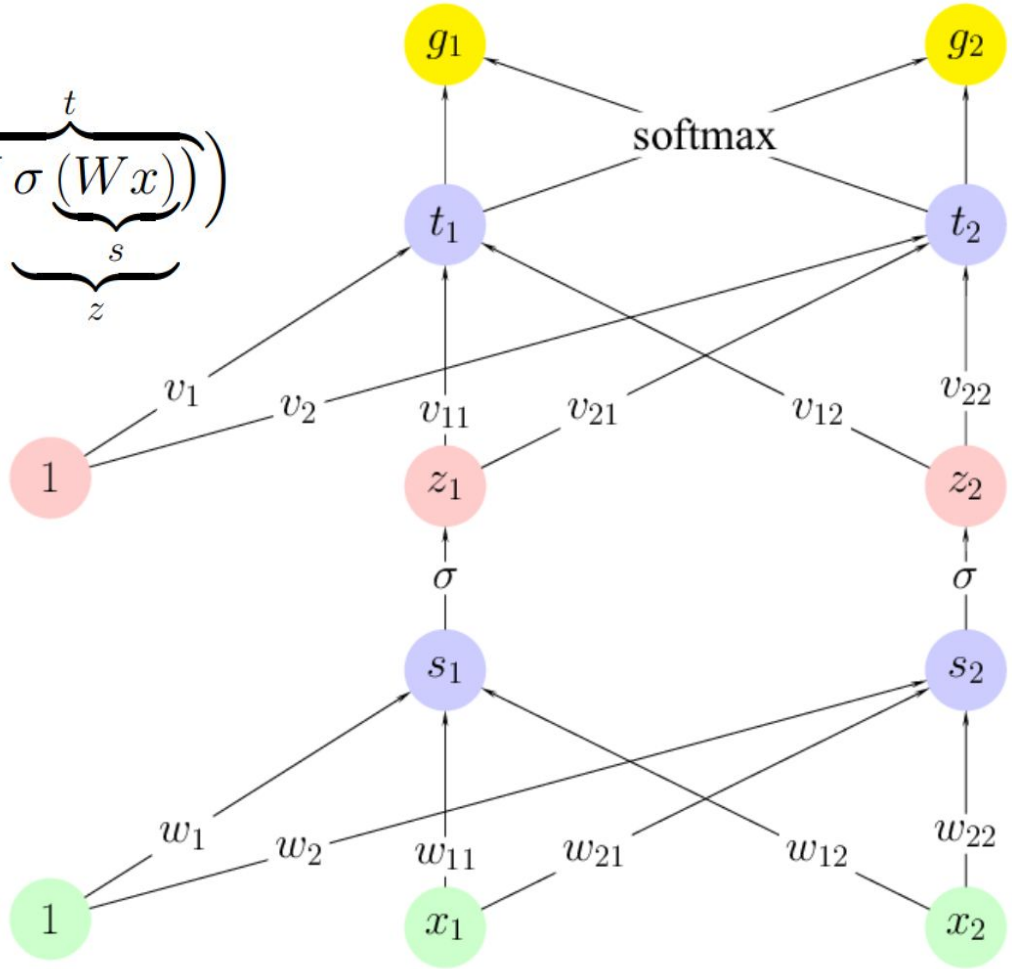
$$g = \text{softmax}\big(V\sigma(Wx)\big)$$

$$R^{(i)} = \text{logloss}(g) = \text{logloss}\Big(\text{softmax}\big(\overbrace{V\,\sigma\,\underbrace{(Wx)}_{s}}^{t}\big)\Big)$$
$$z$$

$$\delta_x = \frac{\partial R^{(i)}}{\partial x} = \overbrace{\underbrace{(g-y)}_{\delta_t}\cdot V\cdot \text{diag}\big(\sigma'(s)\big)}^{\delta_s}\cdot W$$
$$\delta_z$$

$$\frac{\partial R^{(i)}}{\partial W} = \delta_s \cdot x, \qquad \frac{\partial R^{(i)}}{\partial V} = \delta_t \cdot z,$$

$$W \leftarrow W - \rho\frac{\partial R^{(i)}}{\partial W}, \qquad V \leftarrow V - \rho\frac{\partial R^{(i)}}{\partial V},$$

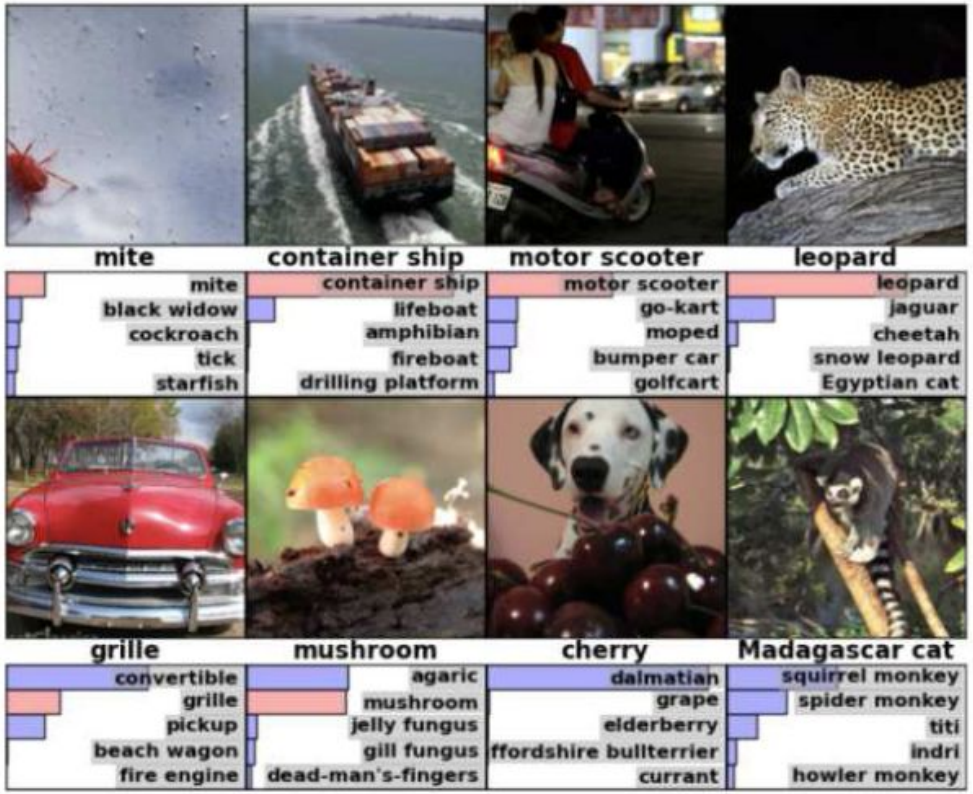


2012 ImageNet и AlexNet

ImageNet ILSVRC-2012
(около 1 млн. изображений, 1000 классов)

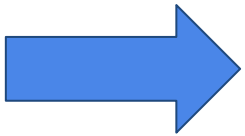
Ошибку удалось понизить с 26% до 15% (сейчас меньше 1%)

A. Krizhevsky, I. Sutskever, **G.E. Hinton**, Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems. 1: 1097–1105, 2012



Глубокое обучение

- Больше данных
- Глубже модели
- Дольше обучение



Выше точность!

Сверточные сети

Линейный фильтр $I*K$ с ядром K:

$$(I * K)_{pq} = \sum_{i=1}^{h} \sum_{j=1}^{w} I_{p+i-1,q+j-1} K_{ij}$$

Например,

$$K = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

0	255	255	255	255	255
0	0	255	255	255	255
0	0	0	255	255	255
0	0	0	0	255	255
0	0	0	0	0	255
0	0	0	0	0	0

0	1	0
1	−4	1
0	1	0

$$(I * K)_{pq} = \sum_{i=1}^{h} \sum_{j=1}^{w} I_{p+i-1,q+j-1} K_{ij}$$

$$K = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

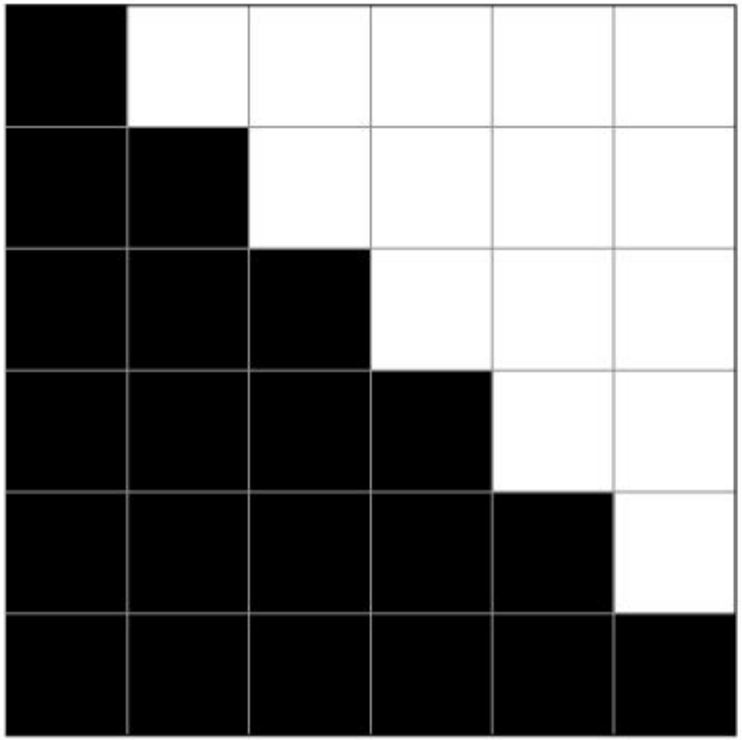
0	255	255	255	255	255
0	0	255	255	255	255
0	0	0	255	255	255
0	0	0	0	255	255
0	0	0	0	0	255
0	0	0	0	0	0

0	1	0
1	−4	1
0	1	0

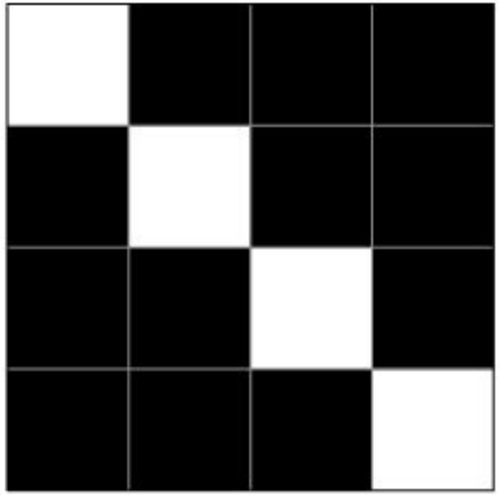
510	−510	0	0
0	510	−510	0
0	0	510	−510
0	0	0	510

$$(I * K)_{pq} = \sum_{i=1}^{h} \sum_{j=1}^{w} I_{p+i-1,q+j-1} K_{ij}$$

$$K = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$



0	1	0
1	−4	1
0	1	0



Линейный фильтр (*свертка*) $I * K$ с ядром K:

$$(I * K)_{pq} = \sum_{i=1}^{h} \sum_{j=1}^{w} I_{p+i-1,\, q+j-1} K_{ij}$$

$$K = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Сверточные сети

Основная идея сверточных сетей (сверточных слоев):
Параметры фильтров будем подбирать с помощью обучения

$$z_{pq} = \sigma\left(\beta_0 + \sum_{i=1}^{h}\sum_{j=1}^{w}\beta_{ij}x_{p+i-1,q+j-1}\right)$$

x_{ij} – узлы (нейроны) одного слоя (например, входного)

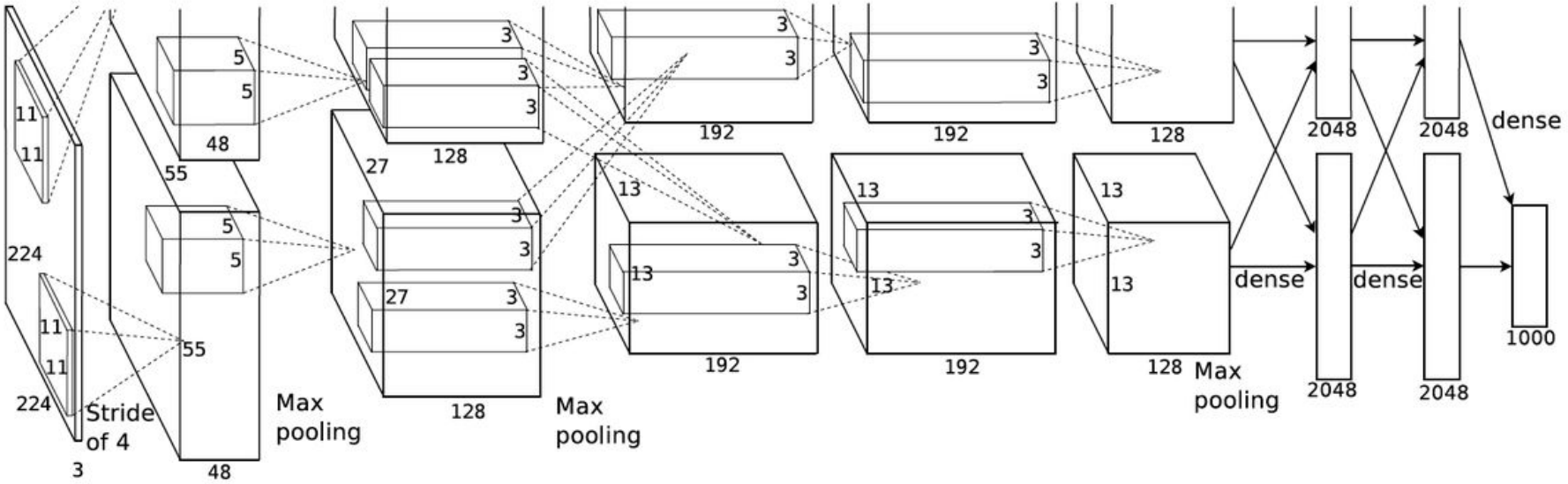
z_{pq} – узлы следующего слоя

Параметры фильтра – это теперь веса нейронной сети.

Отличия от полносвязной сети (полносвязного слоя):

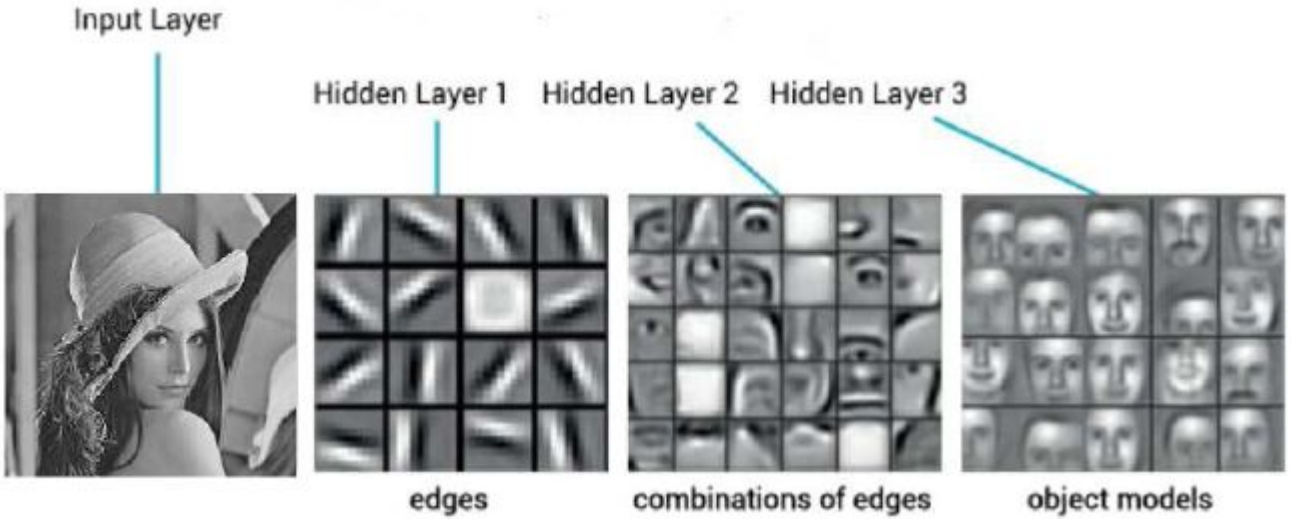
- Нет соединения каждого узла одного слоя со всеми узлами следующего.
- Веса становятся *разделяемыми*.

Alexnet (2012)



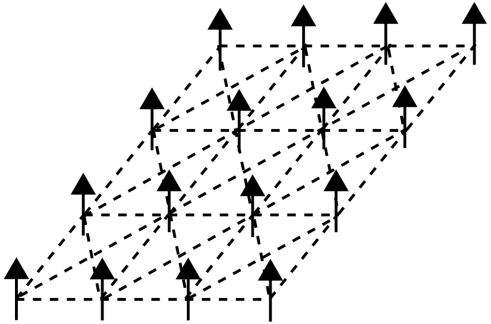
- Сверточные слои (convolutional layers)
- «Выборочные» слои, или слои объединения (subsampling/pooling layers)
- Полносвязные слои (fully connected layers)

Автоматическое выделение признаков

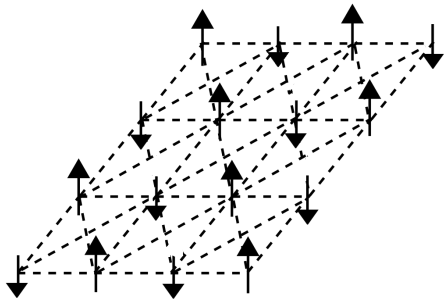


2. 2. Сеть Хопфилда (ассоциативная память) и др.

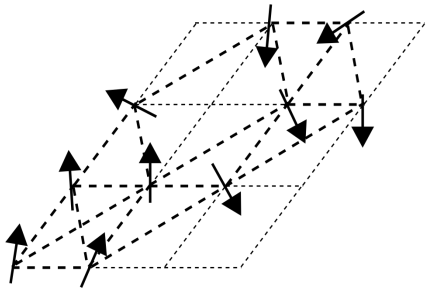
Взаимодействующие магнитные моменты



Ферромагнетики ($M_{ij} > 0$)
– спины стремятся сориентироваться параллельно



Антиферромагнетики ($M_{ij} < 0$)
– спины соседних атомов противоположно направлены



Спиновые стекла
– связи между атомами носят случайный характер

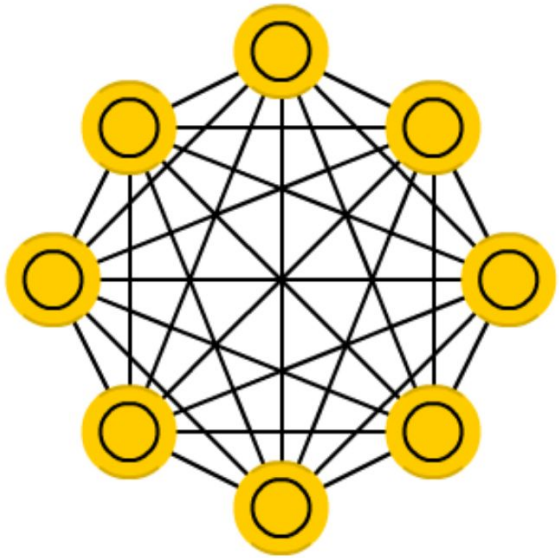
Энергия системы:

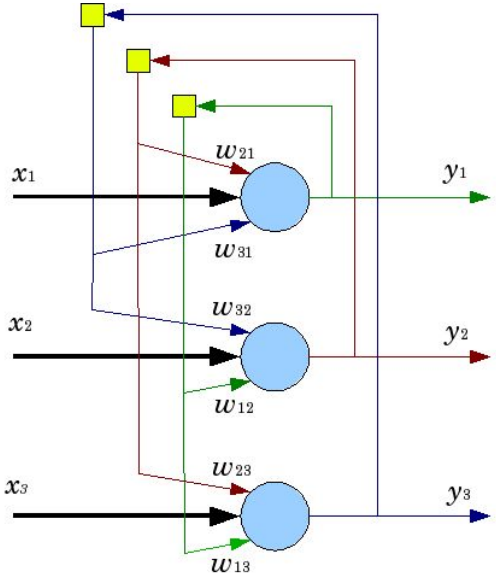
$$E(s) = -\frac{1}{2}\sum_{i,j} M_{ij} s_i s_j = -\frac{1}{2} s^\top M s, \quad s_i \in \{-1, 1\}, \quad M_{ij} = M_{ji}$$

где M_{ij} – коэффициент обменного взаимодействия (Модель Изинга)

Сети Хопфилда подобны спиновым стеклам со множествам аттракторов, которым соответствуют минимумы энергии

Сеть Хопфилда

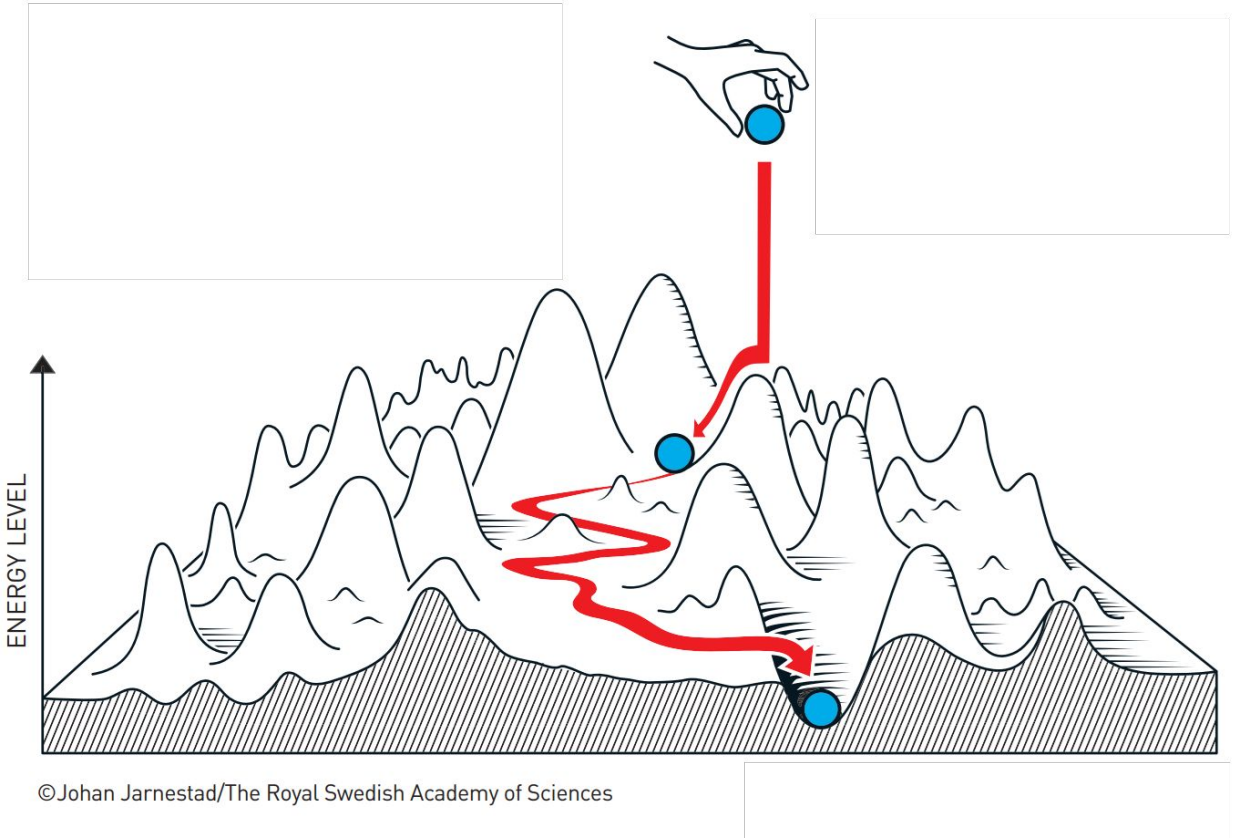




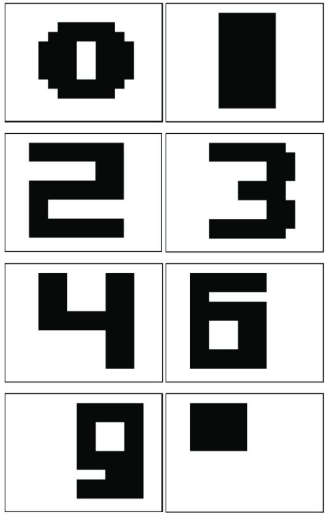
$$y^{(0)} = x, \qquad y^{(t+1)} = \operatorname{sign}(My^{(t)})$$

$$M = (M_{ij})_{d\times d}, \quad M_{ij} = M_{ji}, \quad M_{ij} = 0$$

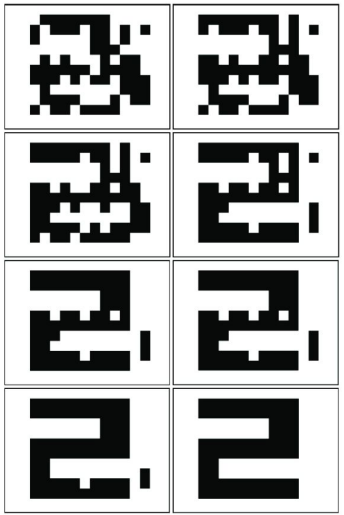
Минимизация энергии



Восстановление шаблона по зашумленному изображению



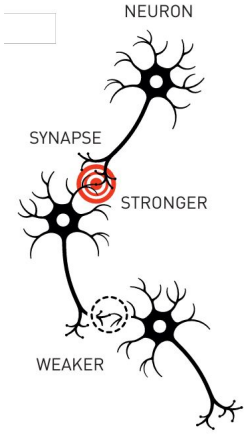
Шаблоны $x^{(i)}$



Примеры

$$y^{(t+1)} = \text{sign}(My^{(t)})$$

$$M = \frac{1}{N}\left(\sum_{i=1}^{N} x^{(i)} x^{(i)\top}\right) - I_N = \frac{1}{N}\mathbf{X}^\top\mathbf{X} - I_N$$



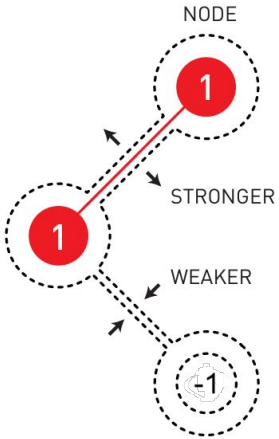


Рис.: Singh, Narotam & Kapoor, Amita. (2015) Cloud Hopfield neural network: Analysis and simulation 10.1109/ICACCI.2015.7275610.

Правило Хэбба: Два взаимодействующих нейрона одного знака усиливают связь между ними. Разных знаков – ослабляют.

Математический взгляд

$$x^{(1)}, x^{(2)}, \ldots, x^{(N)} \in \{-1, 1\}^d$$

$$E(x) = -\frac{1}{2}\sum_{i=1}^{N}\langle x^{(i)}, x\rangle^2 = -\frac{1}{2}x^\top\left(\sum_{i=1}^{N} x^{(i)}x^{(i)\top}\right)x = -\frac{1}{2}x^\top W x \;\to\; \min_x$$

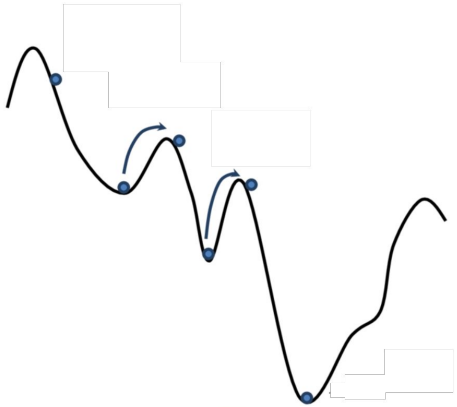
$$M = \frac{1}{N}\left(\sum_{i=1}^{N} x^{(i)}x^{(i)\top}\right) - I_N = \frac{1}{N}\mathbf{X}^\top\mathbf{X} - I_N$$

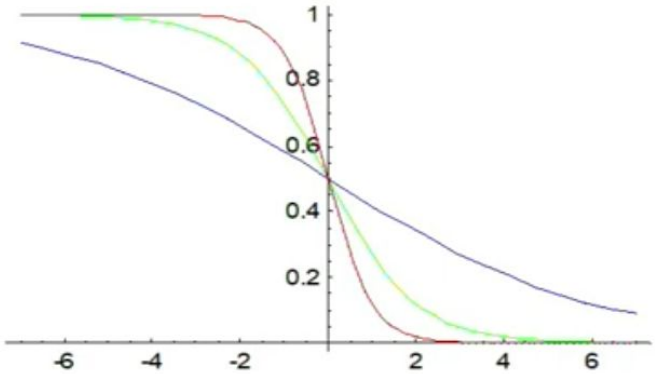
$$M = (M_{ij})_{d\times d}, \quad M_{ij} = M_{ji}, \quad M_{ij} = 0$$

$$y^{(0)} = x$$

$$y^{(t+1)} = \operatorname{sign}(My^{(t)})$$

Машина Больцмана





1. Случайно выбирается нейрон x_j и для него выбирается новое случайное значение из $[-1, 1]$
2. Новое значение принимаем с вероятно

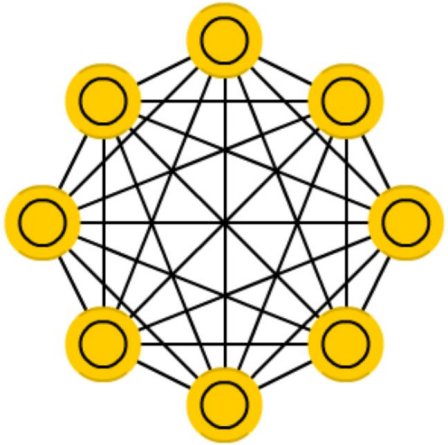
$$P = \begin{cases} 1, & \Delta E_i \leq 0 \\ \frac{1}{1+e^{\frac{\Delta E_i}{T}}}, & \Delta E_i > 0 \end{cases}$$

3. Обновление температуры:

$$T(t) = \frac{T_0}{1 + \ln(t)}$$

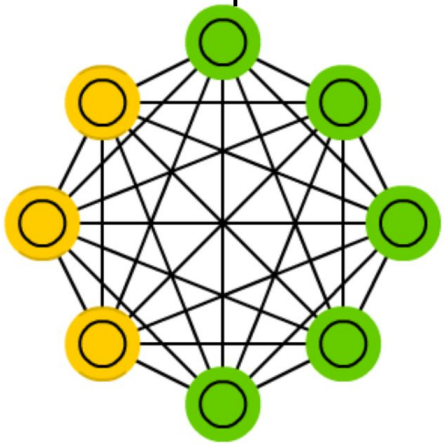
Варианты сети Хопфилда

Сеть Хопфилда



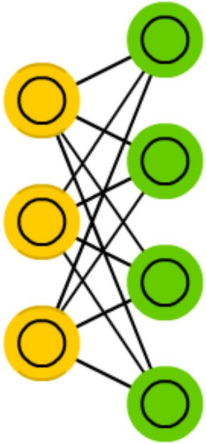
J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities,1982

Машина Больцмана



D.H. Ackley, G.E. Hinton, T.J. Sejnowski, A Learning Algorithm for Boltzmann Machines. 1985

Ограниченная машина Больцмана

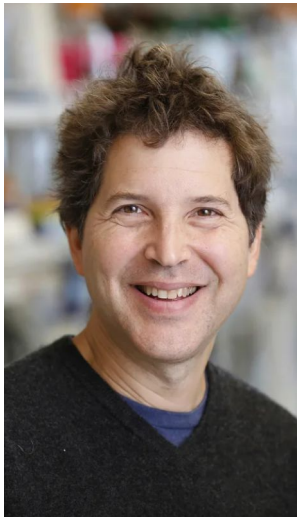


P. Smolensky Harmonium Theory, 1986
G.E. Hinton, R.R.. Salakhutdinov, 2006.

Почему Хопфилд и Хинтон достойны премии по физике?

- «За фундаментальные открытия и изобретения, обеспечивающие машинное обучение с помощью искусственных нейронных сетей»
- Хопфилд и Хинтон – действительно одни из самых крупных ученых в области машинного обучения и ИИ
- Большой вклад в популяризацию нейронных сетей среди ученых
- Машинное обучение и ИИ – новая технология и мощный метод в науке (включая физику)

Нобелевские лауреаты по химии 2024



Дэвид Бейкер
«за вычислительный дизайн белков»

Джон Джампер и Демис Хассабис, Google DeepMind
«за предсказание структуры белков»

Джампер и Хассабис предсказали структуру 200 млн. белков с использованием нейронной сети AlphaFold2, разработанной в DeepMind

Синергия наук

- Стремительное развитие направлений на стыке наук
- Использование методов из других наук – как свежая кровь в исследовании
- Одна из проблем современного ИИ – огромное энергопотребление. Решение – на стыке наук (мемристоры, квантовые алгоритмы, … ?)

Спасибо за внимание!