



Московский государственный университет имени М.В.Ломоносова

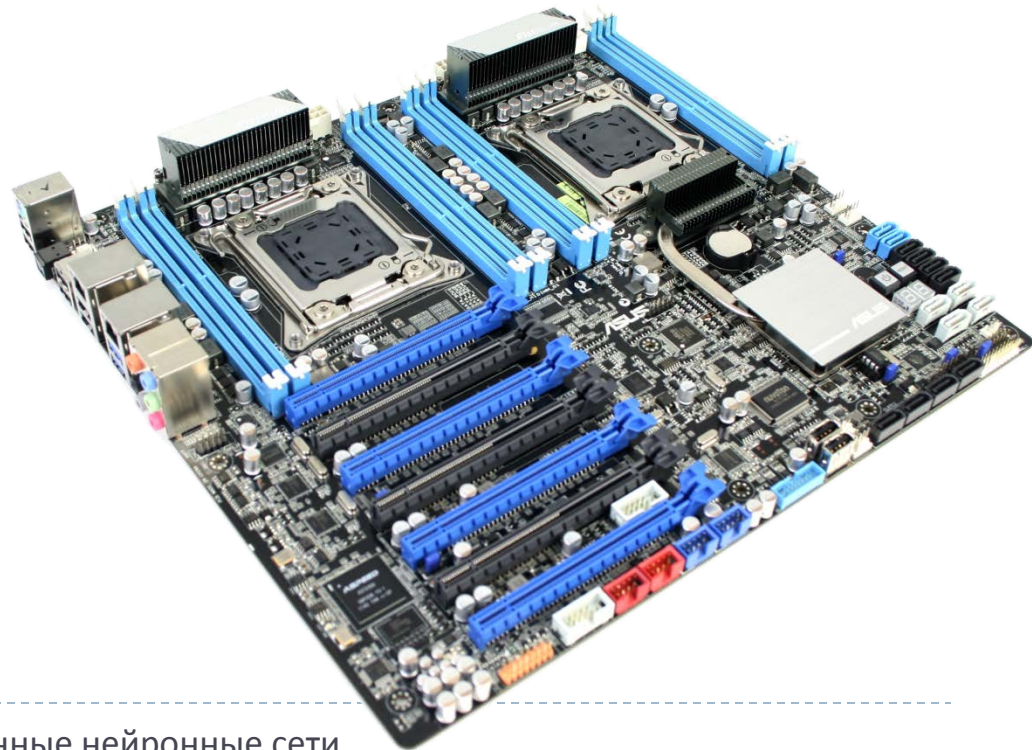
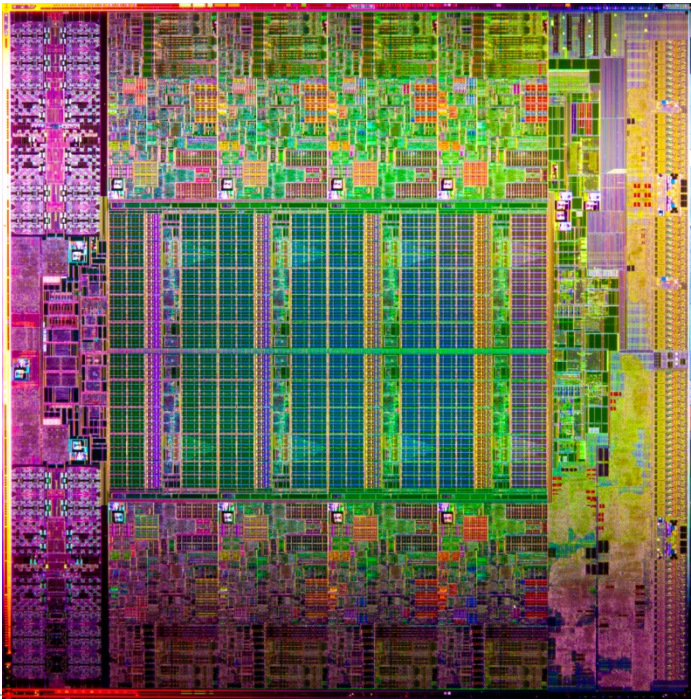
Химический факультет

Искусственные нейронные сети Принципы и приложения

с.н.с., к.х.н. Радченко Евгений Валерьевич
genie@qsar.chem.msu.ru

Современные компьютерные системы

- ▶ Универсальные процессоры
 - ▶ Развитая система команд
 - ▶ Миллиарды транзисторов
 - ▶ Сложные системы поддержки



Современные компьютерные системы

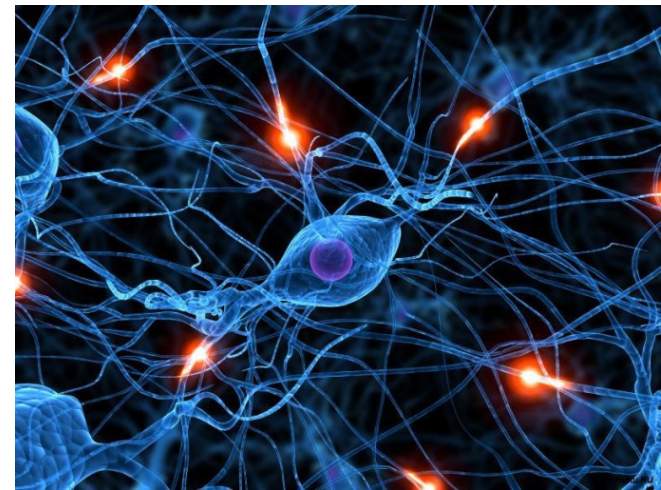
▶ Суперкомпьютеры

- ▶ Параллельная обработка, высокое быстродействие
- ▶ Десятки тысяч узлов на базе универсальных процессоров и других серийных компонентов
- ▶ Значительные требования к размещению, энергоснабжению и охлаждению



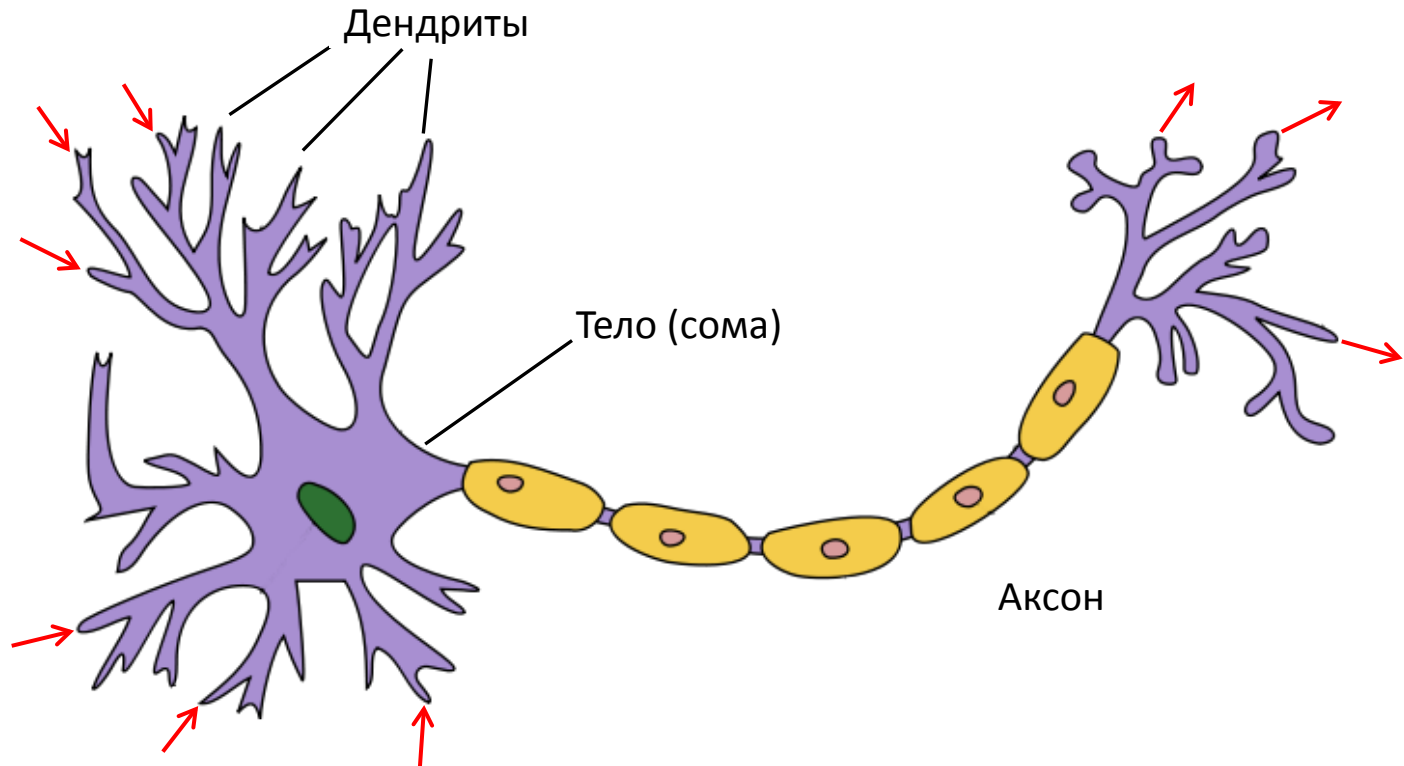
Мозг как устройство обработки информации

- ▶ «Альтернативный» подход
 - ▶ Превосходит доступные сегодня технические решения в таких областях, как зрение, распознавание речи, обучение и др.
 - ▶ Множество ($\sim 10^{11}$) простых однотипных локальных обрабатывающих элементов — **нейронов**
 - ▶ Множество **связей** между нейронами ($\sim 10^4$ у каждого нейрона)
 - ▶ Параллельная обработка
 - ▶ Понимание работы мозга — путь к «искусственному интеллекту»?



Биологический нейрон

- ▶ Получение входных сигналов
- ▶ Объединение и обработка сигналов
- ▶ Передача выходного сигнала

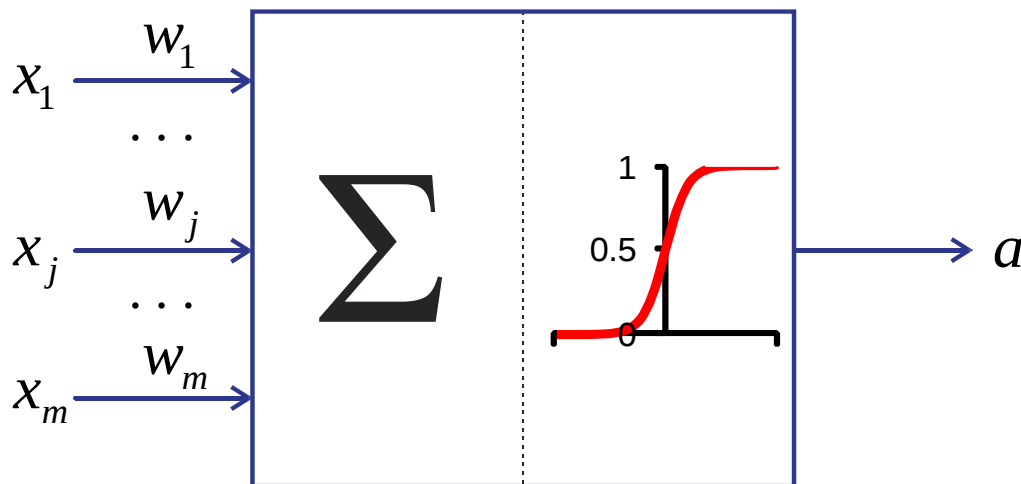


Искусственные нейронные сети

- ▶ [artificial neural networks, ANN]
- ▶ Гибкий общий метод
- ▶ Опирается на упрощенные представления об обработке информации в мозге
- ▶ Простые базовые обрабатывающие элементы – **искусственные нейроны**
 - ▶ Локальная «программа»
- ▶ Соединение нейронов в нейронную сеть
 - ▶ Распределенная обработка информации
- ▶ **Обучение** вместо программирования
- ▶ Простые аппаратные реализации

Искусственный нейрон (перцептрон)

- ▶ Суммирование входных сигналов x
- ▶ Веса соединений (синаптические веса) w
 - ▶ Линейное преобразование (или гиперплоскость)
 - ▶ Взвешенные входы z
- ▶ Как правило, нелинейное преобразование выхода
 - ▶ Функция активации a – обычно сигмоидная



$$z = \sum_m w_j x_j = \mathbf{w}^T \mathbf{x}$$

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

Два типа задач

- ▶ Регрессия (аппроксимация функции)

- ▶ Непрерывная целевая функция y

$$y = f(x) \approx a(x)$$

- ▶ Классификация (распознавание образов)

- ▶ Дискретная целевая функция y – принадлежность к классу (**0 / 1**)
 - ▶ Выход – непрерывная оценка вероятности принадлежности a

$$y = f(x) \approx a(x)$$

- ▶ В обоих случаях можно рассматривать несколько выходов

- ▶ Несколько функций
 - ▶ Несколько классов

Обучение модели как оптимизация

- ▶ Обучающая выборка

$$\{(\mathbf{x}_i, y_i)\}$$

- ▶ Функция штрафа

- ▶ Среднеквадратическая ошибка

$$C = C(w) = \frac{1}{2} \frac{1}{n} \sum_i \|y_i - a(\mathbf{x}_i)\|^2 \rightarrow \min$$

- ▶ Метод градиентного спуска

- ▶ Итеративная процедура минимизации
 - ▶ Скорость обучения (фактор обучения) η

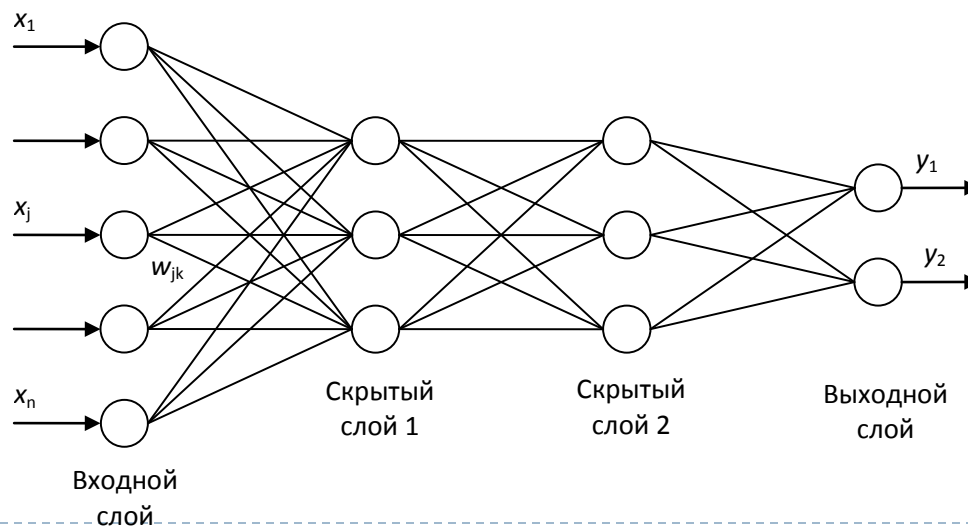
$$\Delta w = -\eta \nabla C \quad \Delta w_j = -\eta \frac{\partial C}{\partial w_j}$$

Многослойные перцептроны

- ▶ Однослойный перцептрон (один уровень преобразования) эквивалентен линейной модели
 - ▶ Линейная регрессия
 - ▶ Линейный дискриминантный анализ
- ▶ Для решения нелинейных задач необходимо соединение нескольких перцептронов
- ▶ **Архитектура нейронной сети** – способ соединения нейронов

Многослойная нейронная сеть прямого распространения

- ▶ Полный набор связей между нейронами соседних слоев
 - ▶ Веса связей w_{jk}
- ▶ Нет других связей между нейронами (внутри слоя или между несоседними слоями)
- ▶ Последовательная передача сигналов с преобразованием
 - ▶ Входной слой – независимые переменные
 - ▶ Выходной слой – прогноз зависимых переменных

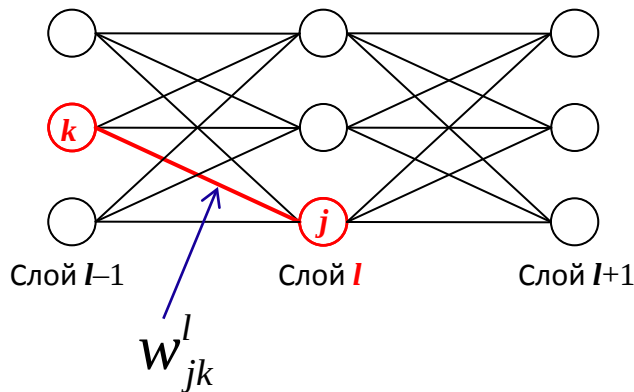


Универсальный аппроксиматор

- ▶ Теорема о представлении Колмогорова-Арнольда
 - ▶ Любую непрерывную функцию нескольких переменных можно представить как конечную суперпозицию функции одной переменной и операций сложения
- ▶ Теорема об универсальной аппроксимации
 - ▶ Нейронная сеть прямого распространения с одним скрытым слоем, включающим конечное число нейронов, может аппроксимировать любую непрерывную функцию
- ▶ Не требуется заранее задавать вид аппроксимируемой функции (например, полином)
 - ▶ Нелинейность «автоматически» воспроизводится скрытым слоем в процессе обучения

Обучение нейронной сети

- ▶ Подбор оптимальных весов связей между нейронами для минимизации функции штрафа (суммы квадратов ошибок)
- ▶ Метод **обратного распространения ошибки** [backpropagation]
 - ▶ Производная сложной функции
 - ▶ Гораздо быстрее по сравнению с численным определением градиента



$$a^l = \sigma(w^l a^{l-1}) = \sigma(z^l)$$

Ошибка нейрона $\delta_j^l = \frac{\partial C}{\partial z_j^l}$

Ошибки нейронов выходного слоя

$$\delta_j^L = \frac{\partial C}{\partial a_j^L} \sigma'(z_j^L) = (a_j^L - y) \sigma'(z_j^L)$$

Или $\delta^L = (a^L - y) \circ \sigma'(z^L)$

Ошибки нейронов предыдущих слоев

$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \circ \sigma'(z^l)$$

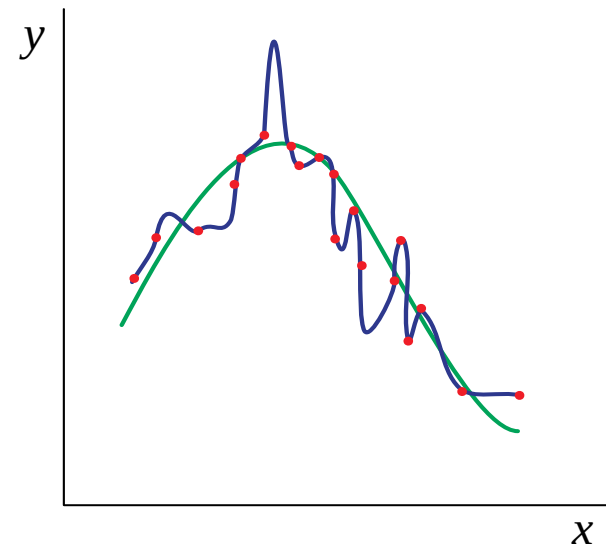
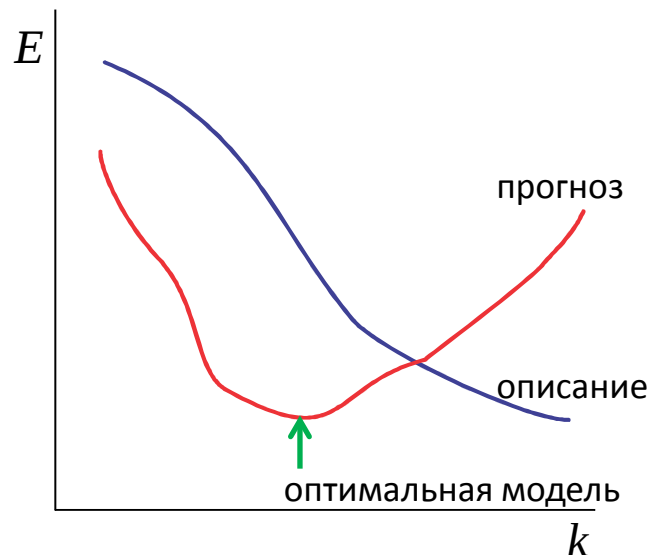
$$\delta^{l+1} \rightarrow \delta^l$$

Производные штрафа по весам связей

$$\frac{\partial C}{\partial w_{jl}^l} = a_k^{l-1} \delta_j^l$$

Переучивание нейронной сети

- ▶ Переучивание нейронной сети – ситуация, когда ошибка прогноза начинает возрастать
 - ▶ Сеть запоминает случайные особенности обучающей выборки за счет неоправданного увеличения сложности модели (для нейросети – увеличение абсолютных значений весов)
 - ▶ Выбор оптимальной модели по предсказательной способности



Глубокие нейронные сети / глубокое обучение

- ▶ [deep neural networks / deep learning]
- ▶ Несколько (много) скрытых слоев
 - ▶ Сеть автоматически обучается все более сложным (высокоуровневым) функциям и/или абстрактным закономерностям
- ▶ Скорость обучения снижается с увеличением числа слоев
 - ▶ «Исчезающие градиенты»
- ▶ Специальные архитектуры сети
 - ▶ Сокращение числа нейронов
 - ▶ Учет внешних ограничений для конкретных задач (например, структура и взаимосвязи входов, симметрия и т.д.)
 - ▶ Например, сверточные сети для распознавания и обработки изображений
- ▶ Метод «грубой силы»
 - ▶ Рост общего уровня производительности вычислительных систем
 - ▶ Специальные аппаратные реализации (прежде всего высокопроизводительные графические процессоры)

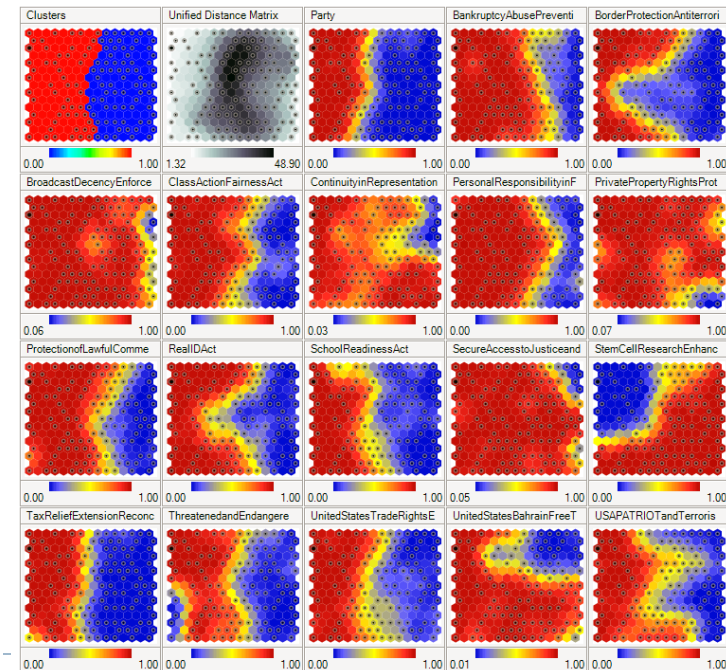
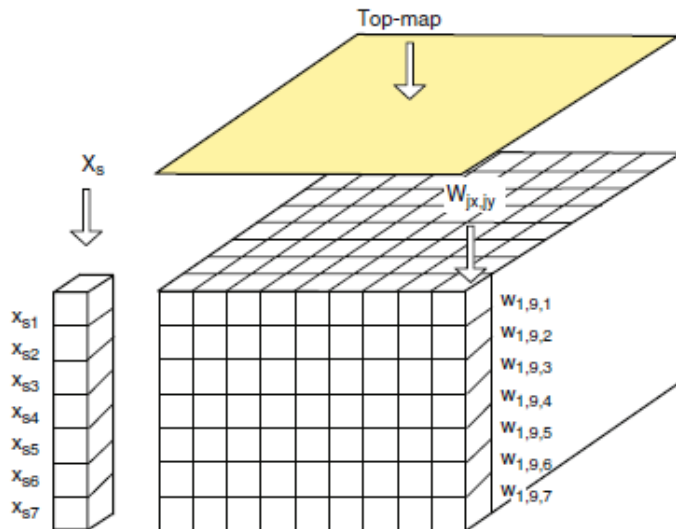
$$\delta^l \sim \boxed{w^{l+1} \sigma'(z^l)} \times \delta^{l+1}$$

Обучение во времени

- ▶ Инкрементное обучение [online learning]
 - ▶ Вариант алгоритма обратного распространения ошибки с последовательным рассмотрением объектов (или их групп)
 - ▶ Иногда более удобен в реализации
 - ▶ Позволяет моделировать меняющиеся закономерности
- ▶ Входы и/или выходы во времени
 - ▶ Распознавание последовательностей (например, речи)
 - ▶ Воспроизведение/прогнозирование последовательностей
 - ▶ Отображение последовательностей
- ▶ Рекуррентные сети: включение дополнительных связей, помимо прямого распространения
 - ▶ Петли, связи с предыдущими слоями
 - ▶ Краткосрочная память для прошлых событий

Самоорганизующиеся карты

- ▶ [self-organizing maps], сети Кохонена
- ▶ Обучение без учителя – «естественная» классификация
- ▶ Одно- или двумерная карта: снижение размерности
 - ▶ Отличается от методов минимизации ошибки
 - ▶ Веса нейронов – векторы в пространстве признаков
 - ▶ В процессе обучения для каждого объекта находится ближайший нейрон
 - ▶ Веса этого (и соседних) нейронов смещаются в сторону объекта
- ▶ В результате сходные объекты отображаются в соседние ячейки
- ▶ Дополнительно – окраска по свойствам

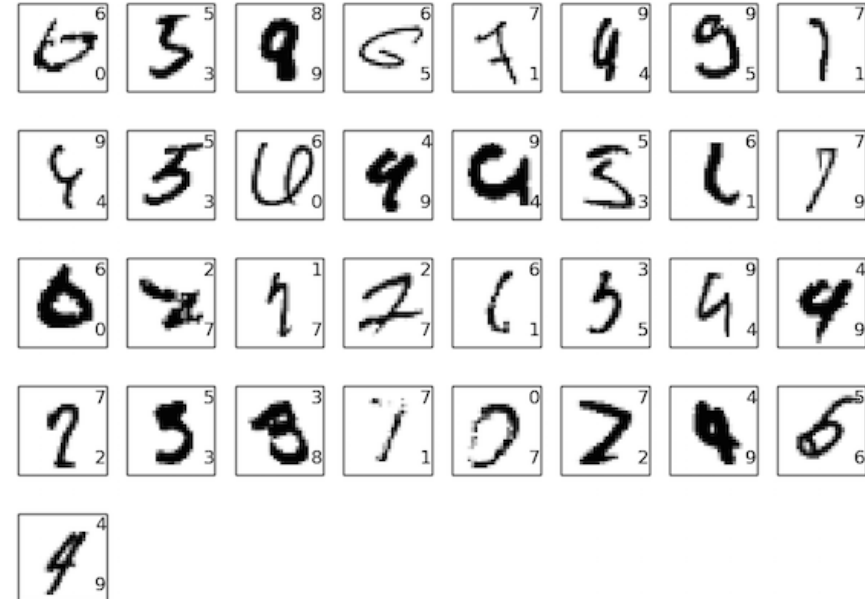


Некоторые приложения

- ▶ Распознавание изображений
- ▶ Обработка изображений
- ▶ Финансовое прогнозирование
- ▶ Моделирования связи структуры химических соединений с их свойствами и биологической активностью
- ▶

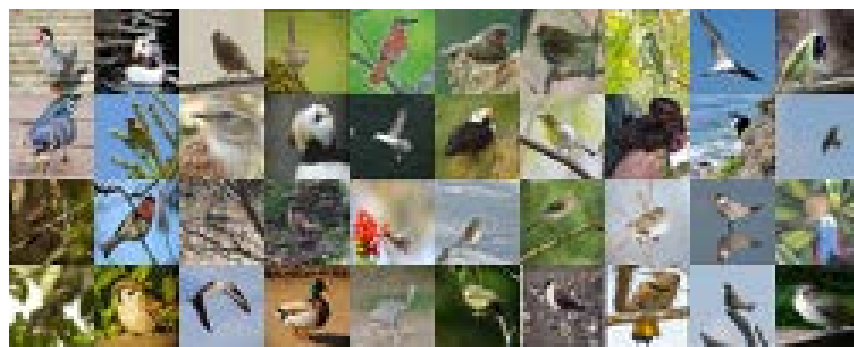
Распознавание изображений

- ▶ Двумерная структура массива входов (пиксели)
 - ▶ Глубокое обучение
 - ▶ Ограничения симметрии
 - ▶ Автоматическое обучение инвариантности к трансляции и вращению
 - ▶ Автоматическое обучение ключевым элементам изображений
-
- ▶ Пример: распознавание рукописных цифр

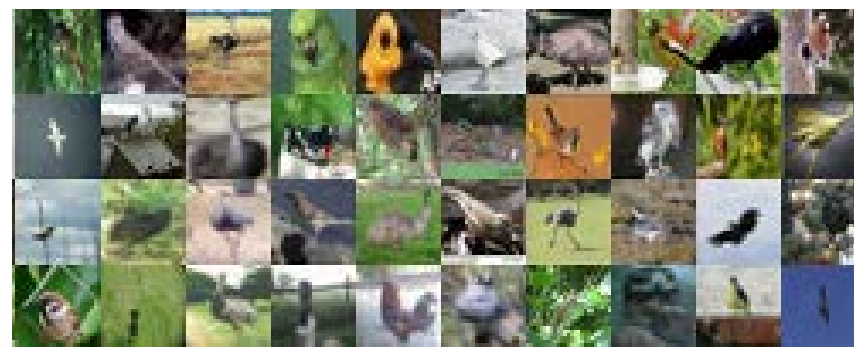


Распознавание изображений

- ▶ Распознавание фотографий
- ▶ Разные виды объектов, разное окружение
- ▶ Пример: птицы



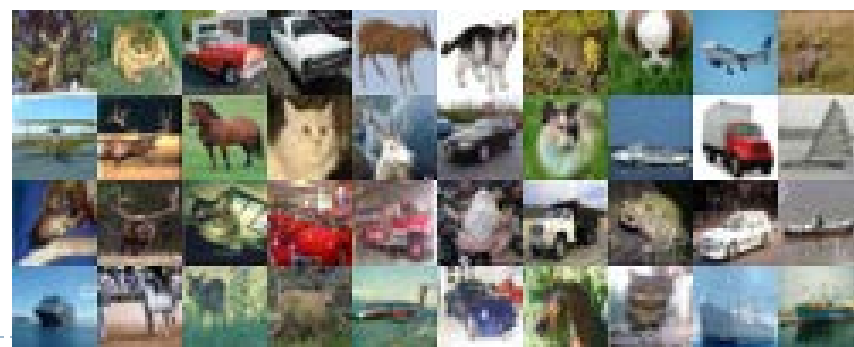
Истинные положительные



Ложные отрицательные



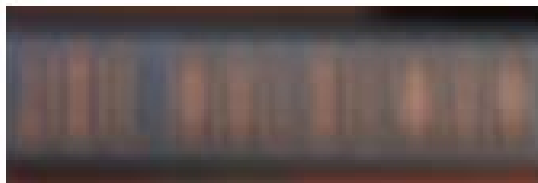
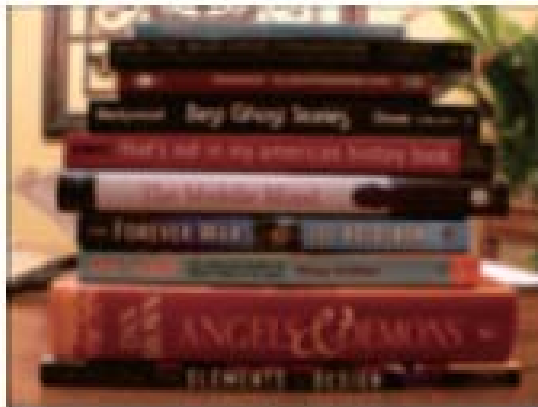
Ложные положительные



Истинные отрицательные

Обработка изображений

- ▶ Двумерная структура массива входов (пиксели)
- ▶ Глубокое обучение
- ▶ Автоматическое обучение ключевым элементам изображений
- ▶ Пример: повышение качества изображений



Обработка изображений

- ▶ Пример: раскраска черно-белых изображений



Черно-белый

Раскраска

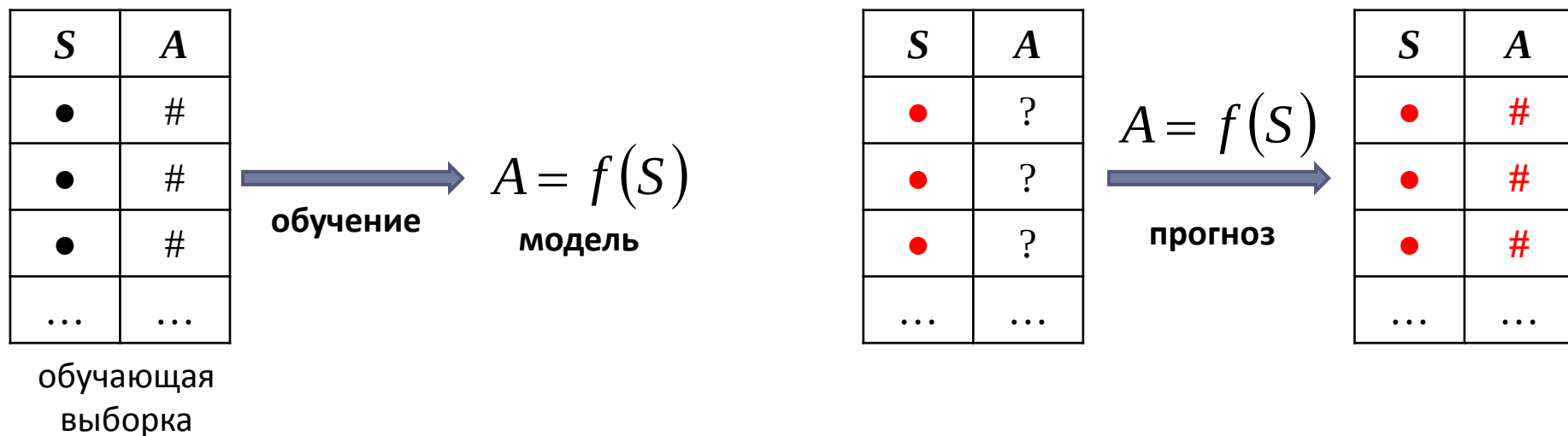
Оригинал

Финансовое прогнозирование

- ▶ Прогнозирование курсов валют и акций
 - ▶ Рекуррентные сети
- ▶ Кредитный скоринг
 - ▶ Классификация
- ▶ Ограничения
 - ▶ Область применимости моделей

Связь структуры химических соединений с их свойствами и биологической активностью

- ▶ Обучающая выборка – набор экспериментальных данных по активности уже исследованных соединений
- ▶ Модель должна обеспечивать **прогнозирование [prediction]** активности для новых соединений
- ▶ Желательна обоснованность и возможность **интерпретации** модели с точки зрения механизма действия соединений и влияющих на активность структурных факторов



Статистический анализ связи структуры и активности

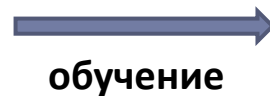
- ▶ Активность как функция описывающих структуру дескрипторов

$$A = f(S) = f(D_1, D_2, D_3, \dots)$$

- ▶ Различные классы функций
- ▶ Различные способы выбора параметров
- ▶ С точки зрения статистического обучения
 - ▶ Активность – **зависимая переменная** y
 - ▶ Дескрипторы – **независимые переменные** $x_1, x_2, x_3, \dots$
 - ▶ Обучающая выборка – набор объектов с известными значениями зависимой и независимых переменных

x_1	x_2	x_3	...	y
#	#	#	...	#
#	#	#	...	#
#	#	#	...	#
...	...	...	...	...

обучающая выборка



$$y = f(x_1, x_2, x_3, \dots)$$

модель

Дескрипторы молекулярной структуры

$$A = f(S) = f(D_1, D_2, D_3, \dots)$$

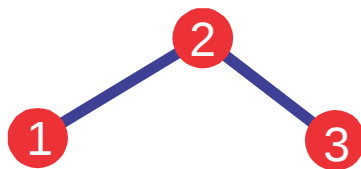
- ▶ Числовые параметры, описывающие те или иные особенности и аспекты структуры соединения
- ▶ Разнообразие доступных дескрипторов почти неограничено
- ▶ Желательна связь с механизмом действия соединений, возможность интерпретации
- ▶ Некоторые типы дескрипторов
 - ▶ Топологические
 - ▶ Физико-химические
 - ▶ Подструктурные
 - ▶ Надструктурные
 - ▶ Дескрипторы трехмерной структуры

Топологические дескрипторы

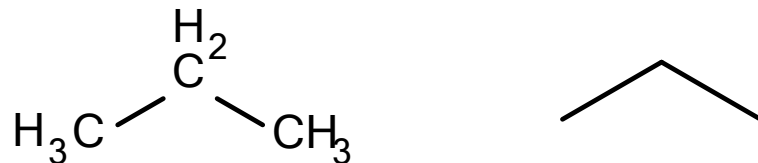
- ▶ Для вычисления достаточно структурной формулы соединения
- ▶ **Инварианты молекулярных графов**
- ▶ **Граф** – математический объект, моделирующий элементы и связи между ними. Образован двумя множествами:
 - ▶ Множество объектов $V = \{ v_i \}$ – **вершины** графа
 - ▶ Множество пар вершин $E = \{ e_i = (v_i, v_j) \}$ – **ребра** графа

Молекулярные графы

► Пример



- Множество вершин $V = \{ 1, 2, 3 \}$
- Множество ребер $E = \{ (1, 2), (2, 3) \}$
- Представляет модель структуры пропана (атомы водорода обычно не рассматривают в явном виде)

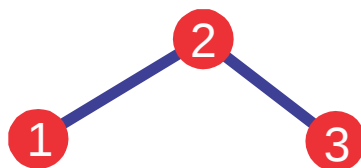


Молекулярные графы

- ▶ Вершины соответствуют атомам
- ▶ Ребра соответствуют связям
- ▶ Вершинам и ребрам могут дополнительно присваиваться атрибуты, отражающие химические особенности атомов и связей
 - ▶ Тип атома, заряд, число атомов водорода, ...
 - ▶ Тип связи, ...

Матрица смежности

- ▶ Симметричная квадратная матрица A размера $N \times N$
- ▶ $A_{ij} = 1$, если атомы i и j связаны



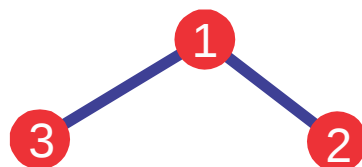
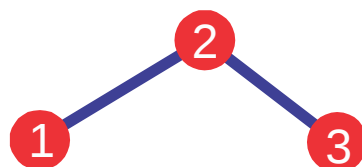
A	1	2	3
1	0	1	0
2	1	0	1
3	0	1	0

- ▶ Дополнительно можно учитывать атрибуты атомов и связей
 - ▶ Представление структур в базах данных – обычно только список связей (ненулевые элементы)

число атомов и связей									
3	2	0	0	0	0	999			
координаты атомов						V2000			
	-1.6927	-0.4439	0.0000	C	0	0	0	0	0
	-0.9782	-0.0314	0.0000	C	0	0	0	0	0
	-0.2638	-0.4439	0.0000	C	0	0	0	0	0
связи						атрибуты атомов			
1	2	1	0	0	0	0	0	0	0
2	3	1	0	0	0	0	0	0	0

Инварианты молекулярных графов

- ▶ Вершины графа можно нумеровать произвольным образом
- ▶ Но граф (и структура) остается тем же самым



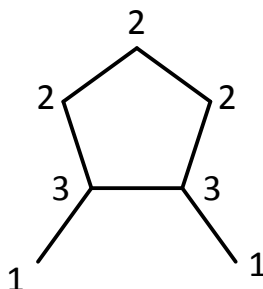
A	1	2	3
1	0	1	0
2	1	0	1
3	0	1	0

A	1	2	3
1	0	1	1
2	1	0	0
3	1	0	0

- ▶ **Инвариант графа** – числовая характеристика, которая не изменяется при перенумерации вершин
 - ▶ Можно использовать как дескрипторы

Индекс Рандича

- ▶ Степень вершины δ – число связанных с ней вершин
 - ▶ Соответствует первичным/вторичным/третичным атомам



- ▶ **Индекс молекулярной связности Рандича [M. Randić] – сумма по связям**

$$\chi = \sum_{\text{связи}} \frac{1}{\sqrt{\delta_i \delta_j}}$$

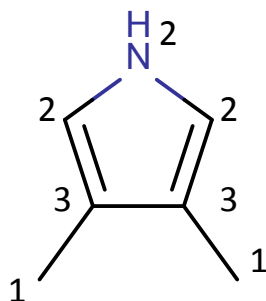
$$\begin{aligned}\chi &= \frac{1}{\sqrt{1 \cdot 3}} + \frac{1}{\sqrt{3 \cdot 2}} + \frac{1}{\sqrt{2 \cdot 2}} + \frac{1}{\sqrt{2 \cdot 2}} + \frac{1}{\sqrt{2 \cdot 3}} + \frac{1}{\sqrt{3 \cdot 3}} + \frac{1}{\sqrt{1 \cdot 3}} = \\ &= \frac{2}{\sqrt{3}} + \frac{2}{\sqrt{6}} + \frac{4}{3} = 3.304\end{aligned}$$

Индекс Рандича

- ▶ Отражает степень разветвленности структуры
- ▶ Коррелирует с температурой кипения углеводородов

$$T_{кип} = 58\chi - 98$$

- ▶ Индекс не учитывает гетероатомы и кратные связи



$$\begin{aligned}\chi &= \frac{1}{\sqrt{1 \cdot 3}} + \frac{1}{\sqrt{3 \cdot 2}} + \frac{1}{\sqrt{2 \cdot 2}} + \frac{1}{\sqrt{2 \cdot 2}} + \frac{1}{\sqrt{2 \cdot 3}} + \frac{1}{\sqrt{3 \cdot 3}} + \frac{1}{\sqrt{1 \cdot 3}} = \\ &= \frac{2}{\sqrt{3}} + \frac{2}{\sqrt{6}} + \frac{4}{3} = 3.304\end{aligned}$$

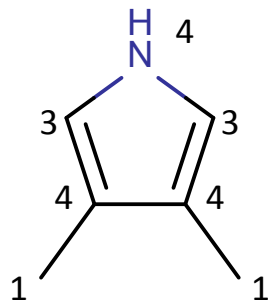
Валентный индекс Рандича

- ▶ Учет гетероатомов и кратных связей
- ▶ Валентная степень вершины

$$\delta^v = \frac{z_v - h}{z - z_v - 1}$$

$$\chi^v = \sum_{\text{связи}} \frac{1}{\sqrt{\delta_i^v \delta_j^v}}$$

- ▶ z – общее число электронов
- ▶ z_v – число валентных электронов
- ▶ h – число присоединенных атомов водорода
- ▶ Знаменатель равен 1 для всех атомов второго периода
- ▶ Для насыщенных атомов углерода совпадает со степенью вершины δ



$$\begin{aligned}\chi &= \frac{1}{\sqrt{1 \cdot 4}} + \frac{1}{\sqrt{3 \cdot 4}} + \frac{1}{\sqrt{3 \cdot 4}} + \frac{1}{\sqrt{3 \cdot 4}} + \frac{1}{\sqrt{3 \cdot 4}} + \frac{1}{\sqrt{4 \cdot 4}} + \frac{1}{\sqrt{1 \cdot 4}} = \\ &= \frac{2}{\sqrt{3}} + \frac{5}{4} = 2.405\end{aligned}$$

Учет более сложных фрагментов

- ▶ **Индексы связности Кира-Холла** [L. Kier, L. Hall]
- ▶ Обобщение индекса Рандича на более сложные фрагменты
- ▶ Порядок индекса – число связей во фрагменте (1 для индекса Рандича)

$$^1\chi = \sum \frac{1}{\sqrt{\delta_i \delta_j}} \quad \text{—}$$

$$^3\chi_p = \sum \frac{1}{\sqrt{\delta_i \delta_j \delta_k \delta_l}} \quad \begin{array}{c} \text{—} \\ \diagup \quad \diagdown \\ \text{—} \end{array}$$

p – path
(путь)

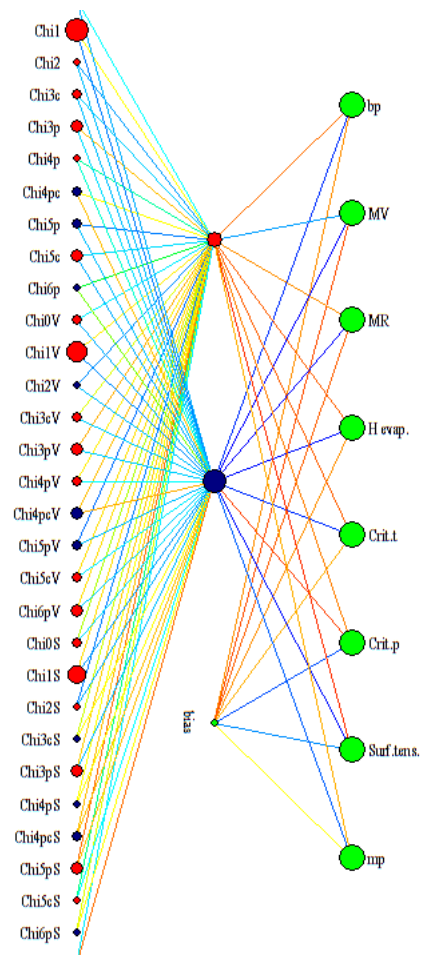
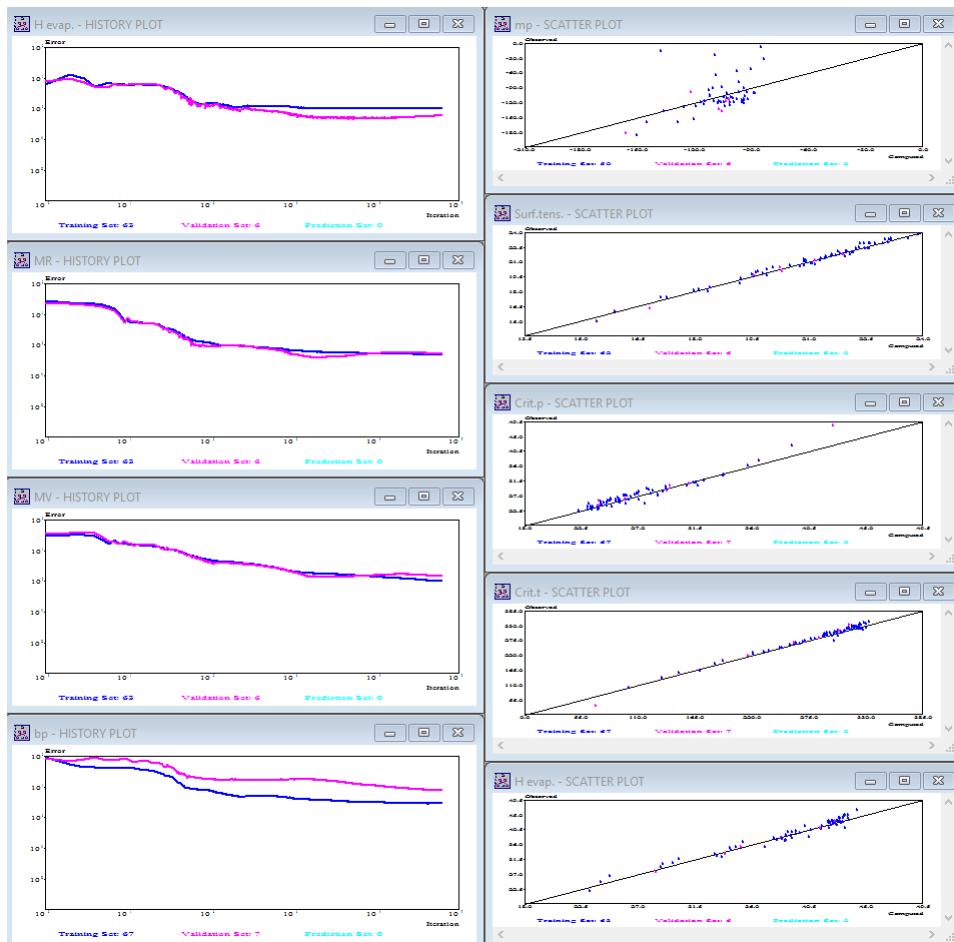
$$^2\chi = \sum \frac{1}{\sqrt{\delta_i \delta_j \delta_k}} \quad \begin{array}{c} \text{—} \\ \diagup \quad \diagdown \end{array}$$

$$^3\chi_c = \sum \frac{1}{\sqrt{\delta_i \delta_j \delta_k \delta_l}} \quad \begin{array}{c} \diagup \quad \diagdown \\ | \end{array}$$

c – cluster
(кластер)

Нейросетевая модель на основе индексов связности

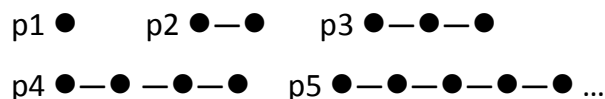
► Различные физико-химические свойства углеводов



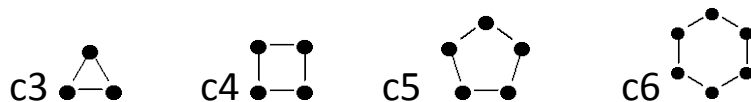
Подструктурные дескрипторы

- ▶ **Подструктурные (фрагментные) дескрипторы** отражают наличие или количество в структуре определенных фрагментов
- ▶ Совокупность фрагментов – «голографический портрет» молекулы
- ▶ Простые базовые фрагменты и классификация атомов

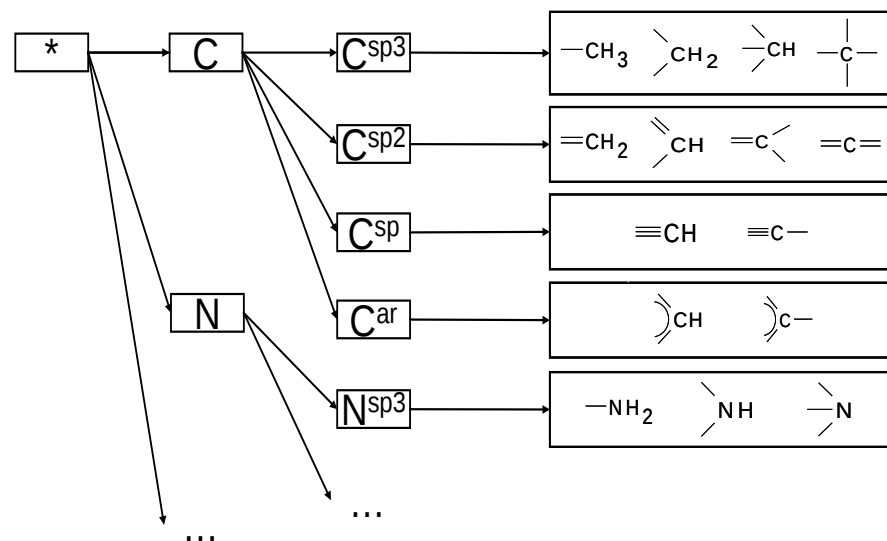
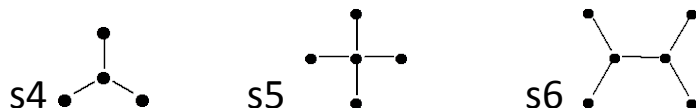
Пути (линейные цепочки)



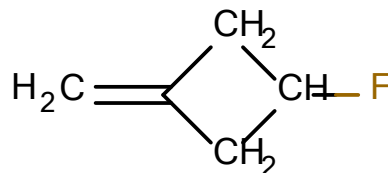
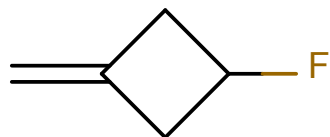
Циклы



Разветвленные фрагменты



Подструктурные дескрипторы



▶ Учитываем

- ▶ Типы атомов
- ▶ Число атомов водорода
- ▶ Типы связей

▶ p1

- ▶ 3x CH₂
- ▶ 1x C
- ▶ 1x CH
- ▶ 1x F

▶ p2

- ▶ 1x CH₂=C
- ▶ 2x CH₂-C
- ▶ 2x CH₂-CH
- ▶ 1x CH-F

▶ p3

- ▶ 2x CH₂=C-CH₂
- ▶ 2x C-CH₂-CH
- ▶ 1x CH₂-C-CH₂
- ▶ 1x CH₂-CH-CH₂
- ▶ 2x CH₂-CH-F

Подструктурные дескрипторы

- ▶ Позволяют учитывать тонкие особенности структуры
 - ▶ Особенно эффективны при построении моделей с помощью **искусственных нейронных сетей**
- ▶ Применимы к широкому ряду разнообразных соединений
- ▶ Легкий прогноз для новых соединений
- ▶ Простая структурная интерпретация
- ▶ Затруднен учет взаимного расположения и влияния фрагментов, экстраполяция на новые классы структур

Нейросетевые модели на основе подструктурных дескрипторов

- ▶ **Физико-химические свойства**
 - ▶ Плотность, температура кипения, энтальпия образования, магнитная восприимчивость...
 - ▶ Реакционная способность
 - ▶ Липофильность
- ▶ **Фармакокинетика, ADME/T**
 - ▶ Всасывание, распределение
 - ▶ Проникновение через гематоэнцефалический барьер
 - ▶ Мутагенность
 - ▶ Токсичность
- ▶ **Активность к конкретным мишеням**
 - ▶ Особенно для широких рядов соединений и разных механизмов действия

Физико-химические свойства

Свойство Параметры модели		Температура кипения, °C		Вязкость $\log(\eta)$, (Pa·s)	Плотность d^{20} , g/cm ³	Давление паров $\log(VP)$, (Pa)	Липофильность $\log P$
Число соединений		509	531	367	803	352	7805
Число дескрипторов		46	54	46	69	56	741
Нейро-сетевая модель	R_{av}	0.9920	0.9960	0.9885	0.9980	0.9981	0.9827
	RMS_{train}	8.7	3.7	0.084	0.021	0.090	0.3233
	RMS_{val}	14.2	4.5	0.104	0.046	0.122	0.3936
	RMS_{pred}	16.6	5.4	0.141	0.051	0.0152	0.3968
Линейная регрессия	RMS_{av}	0.9814	0.9946	0.9794	0.9885	0.9902	0.9702
	RMS_{train}	12.9	4.3	0.111	0.038	0.198	0.4171
	RMS_{val}	16.7	5.0	0.195	0.055	0.248	0.4541
	RMS_{pred}	18.6	5.5	0.212	0.067	0.258	0.4324
Типы фрагментов		p1, p2, p3	p1, p2, p3, p4, p5, p8, c3, c4, c5, c6, c7, c8, c9, s4, b0, b1, b4, b5	p1, p2, p3	p1, p2, p3, p4, p5, c4, c5, s4, s5	p1, p2	px, cx, sx, bx, tx

Фармакокинетические свойства и токсичность (ADMET)

Всасывание

Метаболические изменения

Распределение

Выведение

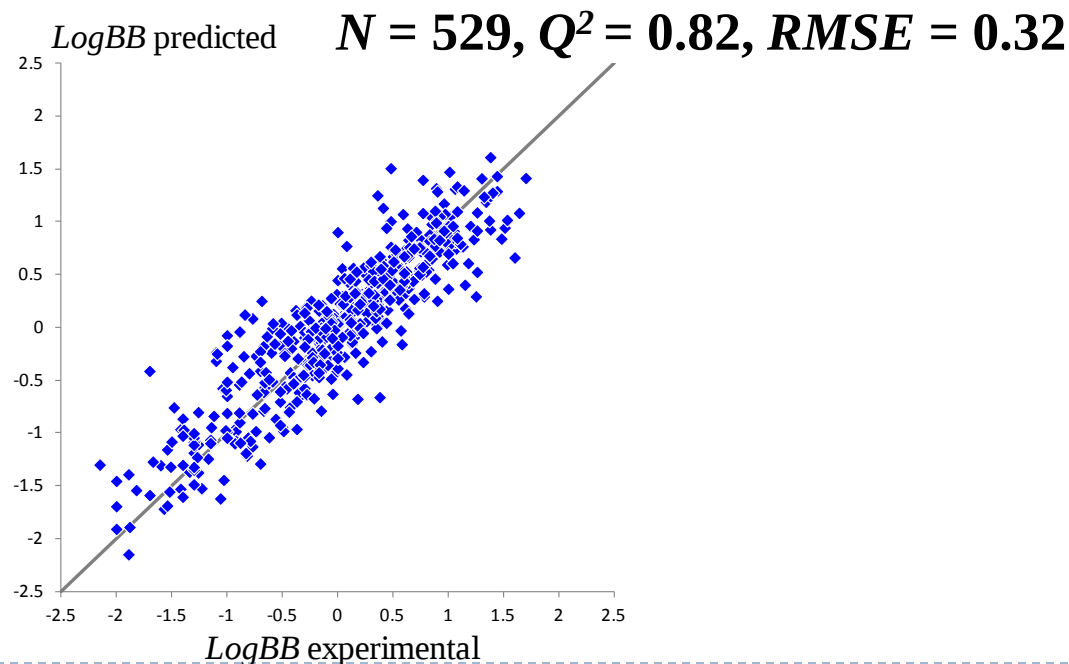
Токсичность

- ▶ Оказывают решающее влияние на **эффективность, фармакологический профиль, способ применения и безопасность действия лекарств**
- ▶ Оптимизация является важной частью поиска и разработки лекарств
- ▶ Возможность прогнозирования для новых структур позволяет значительно повысить скорость и эффективность разработки

Проникновение через гематоэнцефалический барьер

- ▶ Видимо, самая полная и проверенная база данных на основе открытых источников
- ▶ Различные механизмы транспорта не рассматриваются в явном виде
- ▶ Модель сопоставима или превосходит по точности и/или области применимости ранее опубликованные модели

$$\text{LogBB} = \frac{C_{\text{brain}}}{C_{\text{blood}}}$$



Abraham, 2006

$N = 292, R^2 = 0.75, \text{RMSE} = 0.33$

Garg, 2008

$N = 132, Q^2 = 0.79, \text{RMSE} = 0.33$

Muehlbacher, 2011

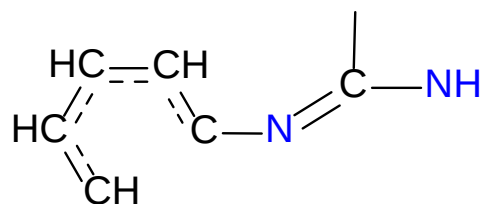
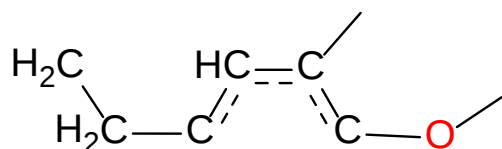
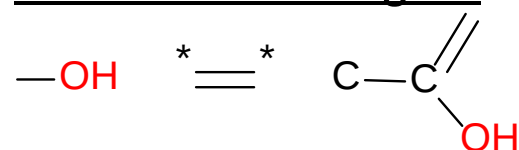
$N = 352, Q^2 = 0.55$

Проникновение через гематоэнцефалический барьер

- ▶ Относительный отклик прогноза по ансамблю нейросетей
- ▶ Разумная интерпретация основных вкладов фрагментов

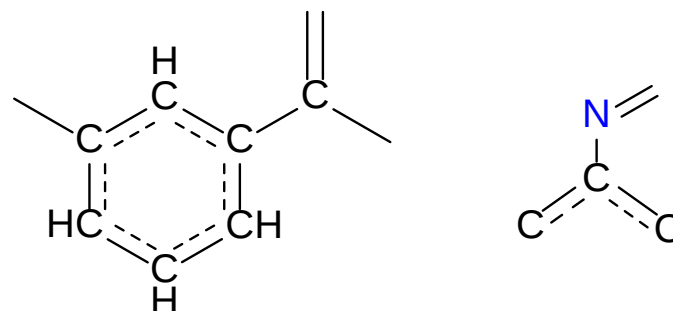
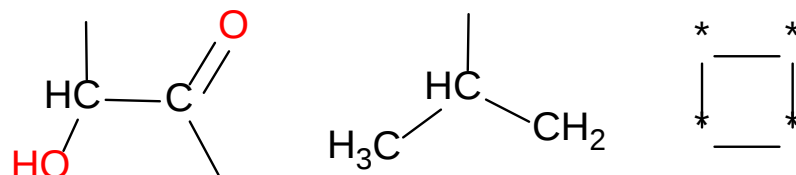
$$RR_i = \text{avg} \left(\frac{\partial y}{\partial x_i} \right) \times \frac{\text{range}(x_i)}{\text{range}(y)}$$

Уменьшение LogBB



Полярные группы

Увеличение LogBB



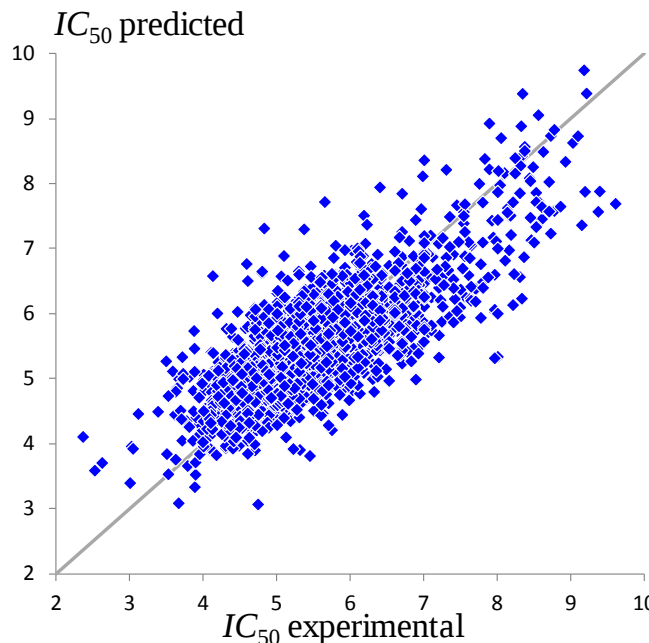
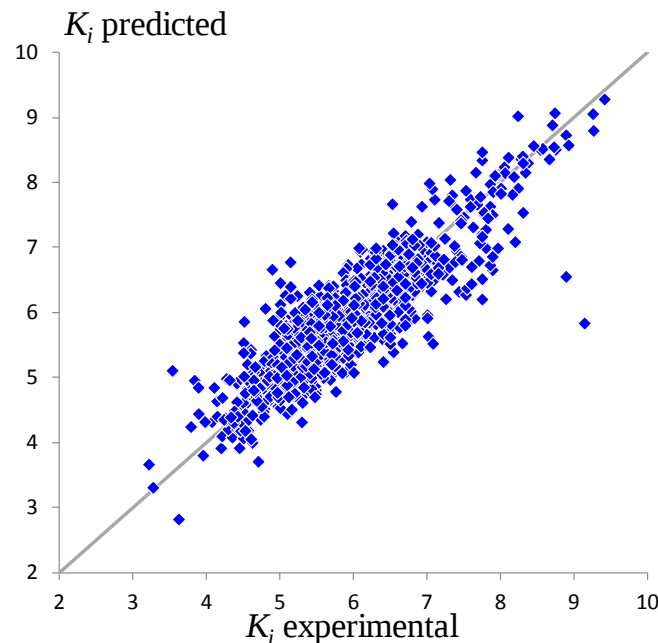
Гидрофобные группы, активный транспорт
замещенных монокарбоновых кислот

Риск hERG-опосредованной кардиотоксичности

- ▶ Калиевый канал кардиомиоцитов hERG – одна из важнейших антимишеней при разработке лекарств
- ▶ Данные измерений аффинности (K_i) и функциональных тестов (IC_{50})
- ▶ Модель сопоставима или превосходит по точности и/или области применимости ранее опубликованные модели

**pK_i : $N = 1000$,
 $Q^2 = 0.77$, $RMSE = 0.45$**

**pIC_{50} : $N = 2886$,
 $Q^2 = 0.60$, $RMSE = 0.55$**



Gavaghan, 2007 (pIC_{50})
 $N = 1312$, $Q^2 = 0.48$, $RMSE = 0.49$

Seierstad, Agrafiotis, 2006 (pIC_{50})
 $N = 439$, $Q^2 = 0.67$, $RMSE = 0.02$

Интегрированная веб-платформа для прогнозирования ADMET-свойств

<http://qsar.chem.msu.ru/admet/>

ADMET Prediction Service

This is an integrated online service for ADMET properties prediction developed at the [Laboratory of Medicinal Chemistry](#) of Department of Chemistry, Lomonosov Moscow State University.

Welcome to the online ADMET prediction service

Please draw a structure (click the "i" icon for drawing help) and then press the Calculate button.

If you have any questions, problems, or suggestions, please do not hesitate to email us
[Dr. Eugene V. Radchenko](#)
[Dr. Vladimir A. Palyulin](#)

SMILES conversion provided by OpenChemLib

Prediction Results

Blood-brain barrier permeability (LogBB)

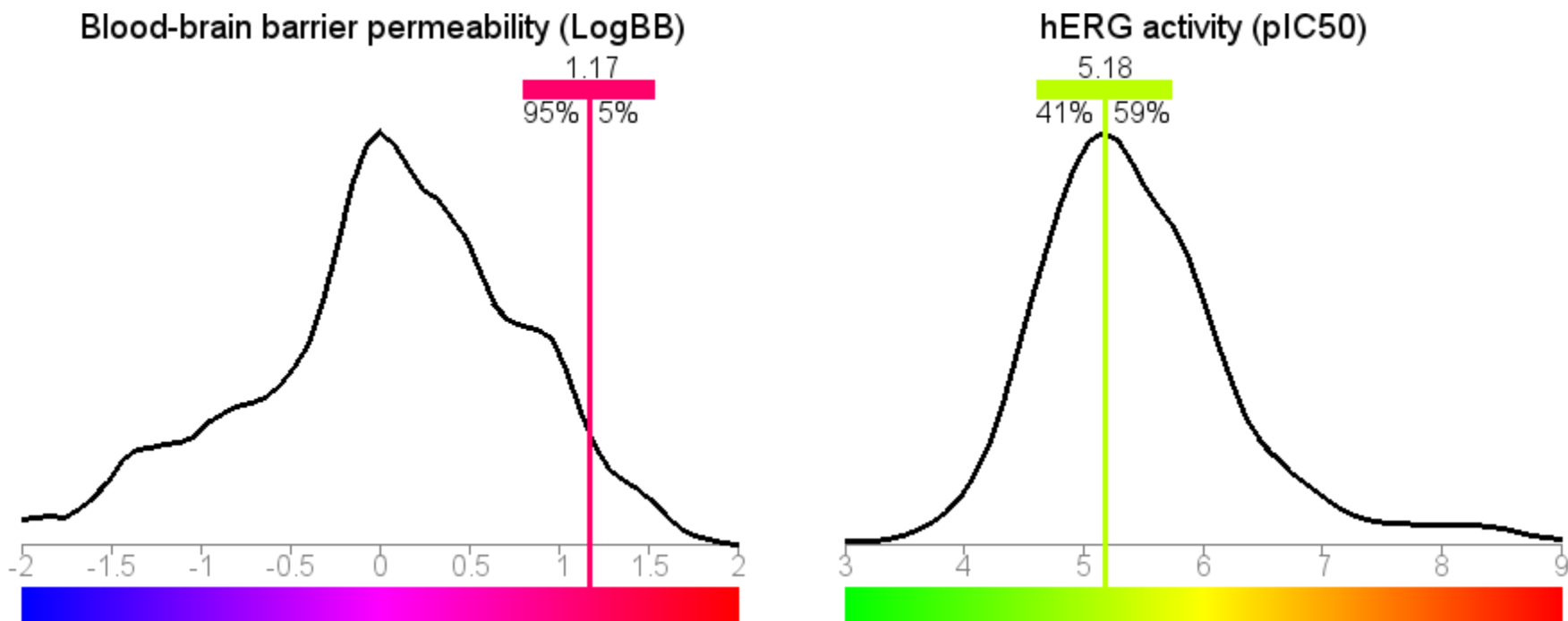
-1.48
3% 97%

Human intestinal absorption (HIA%)

85.37
49% 51%

Интегрированная веб-платформа для прогнозирования ADMEТ-свойств

- ▶ Прогнозируемое значение
- ▶ Цветовая шкала (**хороший**/**плохой** или **низкий**/**высокий**) и место соединения
- ▶ Распределение свойства для известных соединений (обучающей выборки)
- ▶ Доля соединений ниже/выше прогнозируемого значения



Перспективы развития нейросетевых подходов

- ▶ Совершенствование архитектур нейронных сетей и алгоритмов обучения
- ▶ Высокопроизводительные реализации, прежде всего на основе графических процессоров
- ▶ Новые подходы к интерпретации нейросетевых моделей с учетом специфики прикладных задач
- ▶ На пути к искусственному интеллекту?

Благодарности

- ▶ И.И.Баскин
- ▶ В.А.Палюлин
- ▶ Н.С.Зефиров
- ▶ П.В.Карпов
- ▶ А.С.Дябина
- ▶ А.Я.Сафаньяев
- ▶ Ю.А.Рулев

Некоторые обзорные публикации

- ▶ Artificial Neural Networks: Methods and Applications, ed. by D.J.Livingstone, [Springer, 2008](#)
- ▶ LeCun Y., Bengio Y., Hinton G. Deep learning, Nature, 2015, 521, 436–444, [DOI: 10.1038/nature14539](#)
- ▶ Баскин И.И., Палюлин В.А., Зефиров Н.С. Многослойные персептроны в исследовании зависимостей «структура–свойство» для органических соединений, [Росс. хим. журн., 2006, 50\(2\), 86–96](#)
- ▶ Гальберштам Н.М., Баскин И.И., Палюлин В.А., Зефиров Н.С. Нейронные сети как метод поиска зависимостей структура–свойство органических соединений, [Усп. хим., 2003, 72\(7\), 706-727](#)
- ▶ Baskin I.I., Winkler D., Tetko I.V. A renaissance of neural networks in drug discovery, Exp. Opin. Drug Discov., 2016, 11(8), 785–795, [DOI: 10.1080/17460441.2016.1201262](#)

Спасибо за внимание!

