

Лекция 2.

Алгоритмы машинного обучения

Юрий Яровиков

План лекции

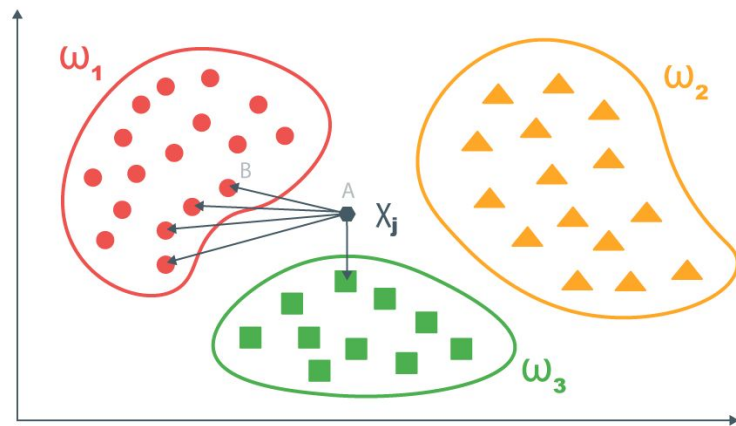
- Метрические алгоритмы
- Логические алгоритмы
- Нейронные сети

Метрические алгоритмы

Метрический алгоритм — алгоритм, опирающийся на геометрическую структуру данных в пространстве объектов.

Алгоритм k ближайших соседей:

- Хотим предсказать класс объекта x
- Вычисляем $f_1(x), f_2(x), \dots, f_n(x)$
- Находим k ближайших объектов из обучающей выборки
- Предсказание = самый популярный класс среди соседей



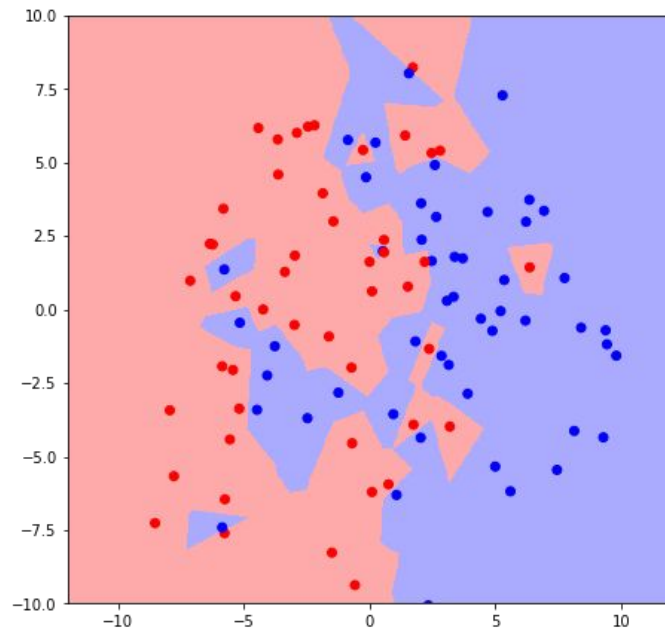
Алгоритм k ближайших соседей

Преимущества алгоритма:

- быстрый и лёгкий в понимании
- легко обобщается для задачи регрессии (как?)

Недостатки алгоритма:

- необходимо хранить всю обучающую выборку
- неустойчив к масштабу признаков



алгоритм одного ближайшего соседа

Логические алгоритмы классификации

Логический алгоритм — алгоритм, использующий логические закономерности в данных.

Примеры простых решающих правил:

- Если в анкете указан домашний телефон и зарплата клиента $> \$2000$ и размер кредита $< \$5000$, то кредит выдать.
- Если возраст пациента > 60 и пациент ранее перенёс инфаркт, то операцию не делать.

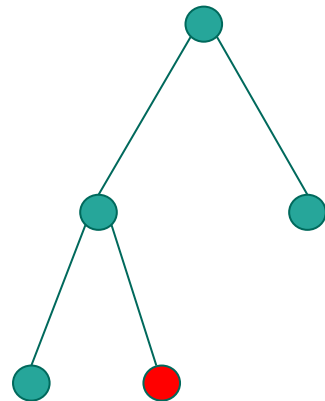
Решающие деревья

- В каждой вершине дерева находится вопрос
- В зависимости от ответа на вопрос, алгоритм направляется в нужную ветвь дерева
- Листы дерева соответствуют решению алгоритма



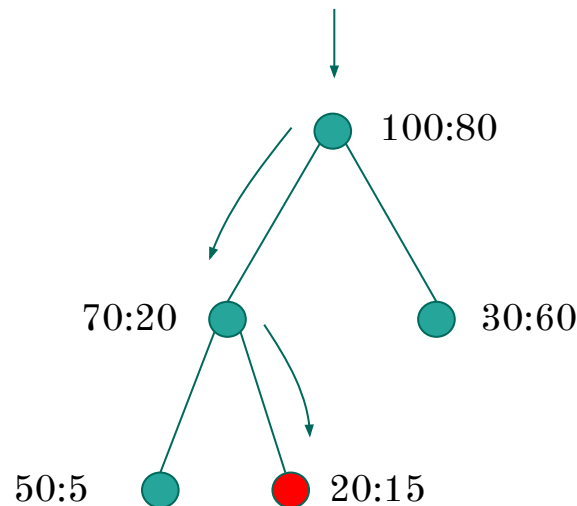
Обучение решающего дерева

- Находимся в красной вершине



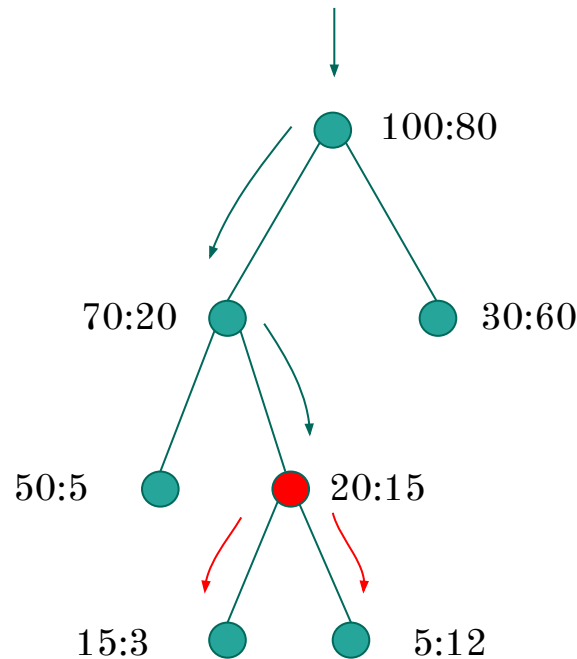
Обучение решающего дерева

- Находимся в красной вершине
- До красной вершины дошла часть объектов обучающей выборки



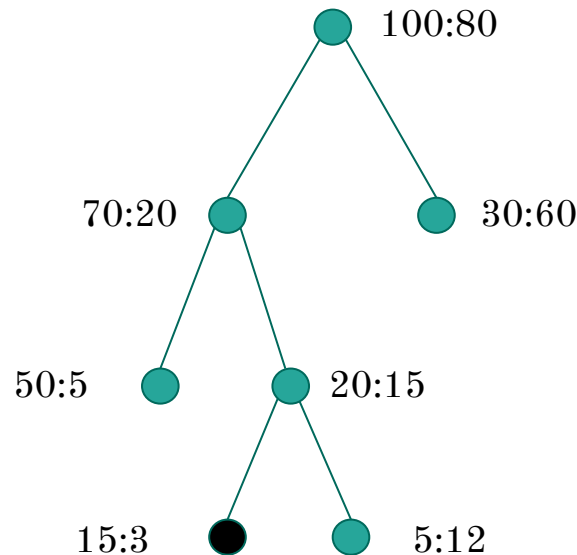
Обучение решающего дерева

- Находимся в красной вершине
- До красной вершины дошла часть объектов обучающей выборки
- Находим решающее правило так, чтобы объекты, дошедшие до красной вершины, хорошо разделялись по искомым классам



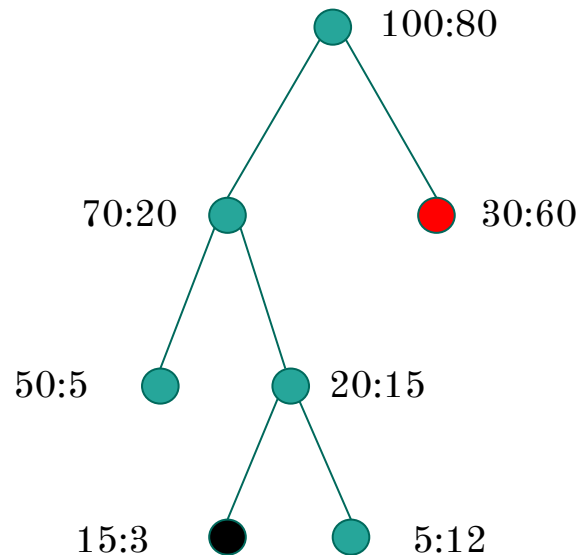
Обучение решающего дерева

- Находимся в красной вершине
- До красной вершины дошла часть объектов обучающей выборки
- Находим решающее правило так, чтобы объекты, дошедшие до красной вершины, хорошо разделялись по искомым классам
- Одна из нижних вершин стала терминальной



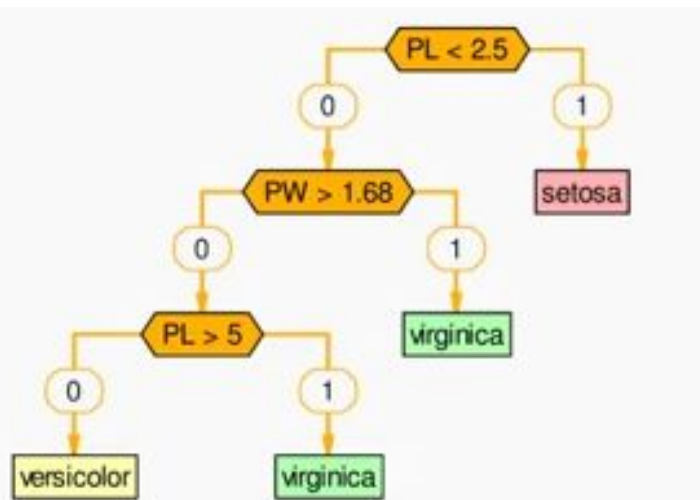
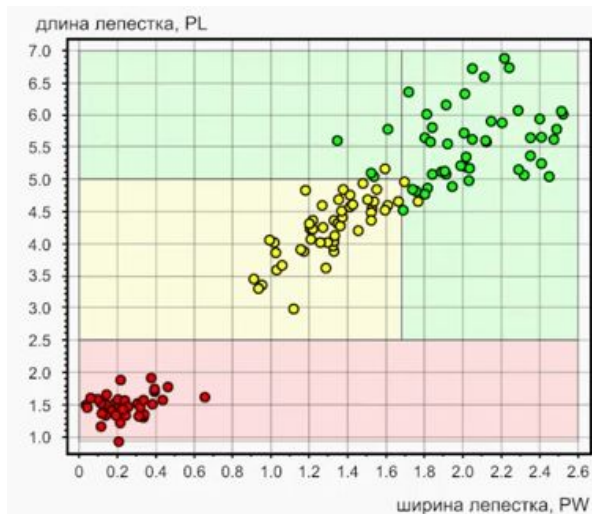
Обучение решающего дерева

- Находимся в красной вершине
- До красной вершины дошла часть объектов обучающей выборки
- Находим решающее правило так, чтобы объекты, дошедшие до красной вершины, хорошо разделялись по искомым классам
- Одна из нижних вершин стала терминальной
- Повторяем с другой вершиной



Решающее дерево на примере Ирисов Фишера

Задача Фишера о классификации ирисов на три класса. В выборке по 50 объектов каждого класса, у каждого объекта 4 признака



В осях двух самых информативных признаков два класса разделились без ошибок, на третьем три ошибки.

Решающий лес

- Построим совокупность решающих деревьев, в каждое из них внеся элемент случайности
- Каждое дерево приняло решение о классификации
- Будем принимать окончательное решение обычным голосованием
- Выходит невероятно эффективно!

Резюме

Преимущества решающих деревьев:

- интерпретируемость
- допускаются разнотипные данные
- возможность обхода пропусков

Недостатки решающих деревьев:

- склонны к переобучению
- фрагментация

Случайный лес — хороший способ устранить недостатки

2006 год: появление проекта Asirra

starm




Please click on all the images that show cats:



Computer Vision: 60%

2014 год: соревнование Dogs VS Cats



Dogs vs. Cats

Create an algorithm to distinguish dogs from cats
215 teams · 5 years ago





[Overview](#) [Data](#) [Kernels](#) [Discussion](#) [Leaderboard](#) [Rules](#)

[Public Leaderboard](#) [Private Leaderboard](#)

The private leaderboard is calculated with approximately 70% of the test data.
This competition has completed. This leaderboard reflects the final standings.

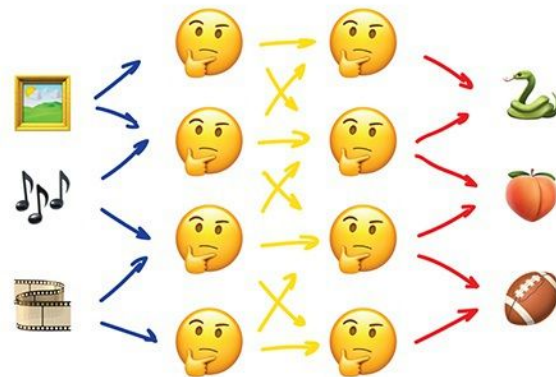
[Refresh](#)

■ Gold ■ Silver ■ Bronze

#	Δpub	Team Name	Kernel	Team Members	Score 🏆	Entries	Last
1	—	Pierre Sermanet			<u>0.98914</u>	5	5y
2	▲ 4	orchid			0.98308	17	5y
3	—	Owen			0.98171	15	5y
4	—	Paul Covington			0.98171	3	5y

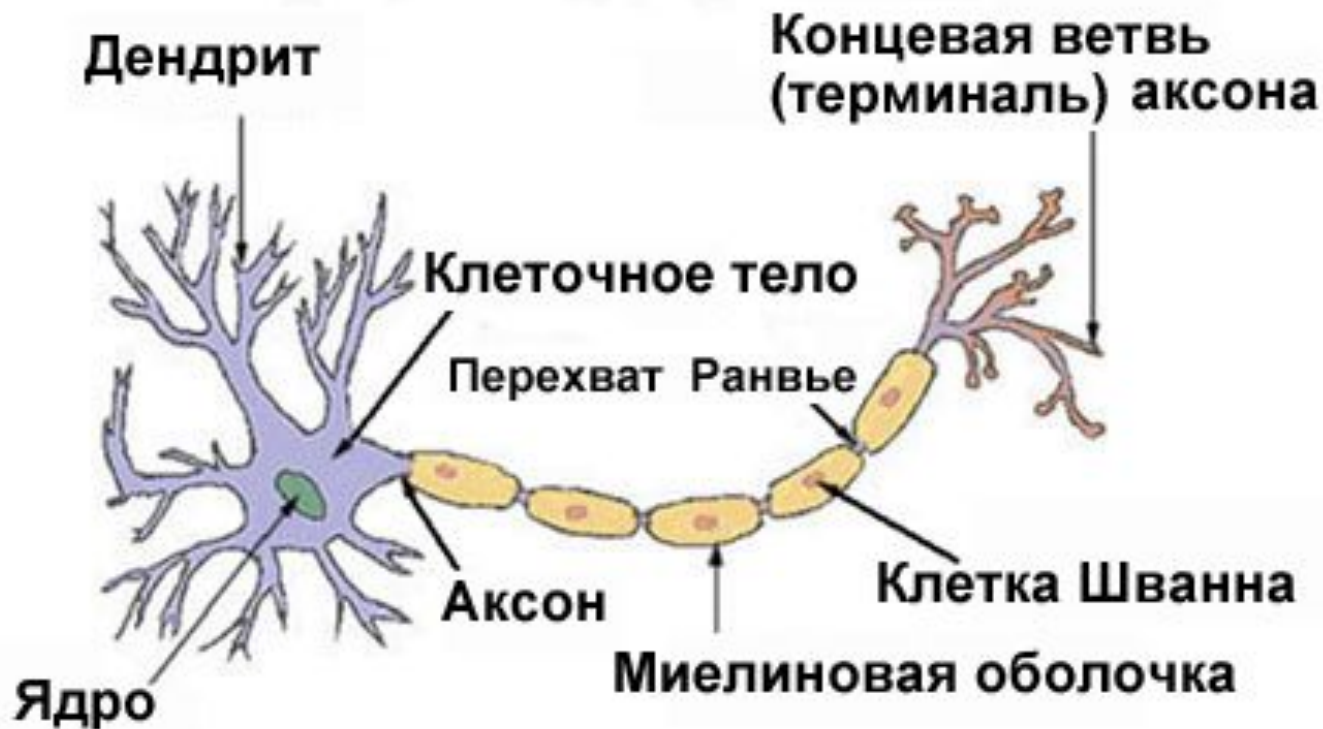
Нейронные сети

- Самая быстроразвивающаяся и многообещающая ветвь машинного обучения
- Пригодна для обработки большого количества данных
- Берёт на себя процесс feature engineering
- Способна к “переносу знания”

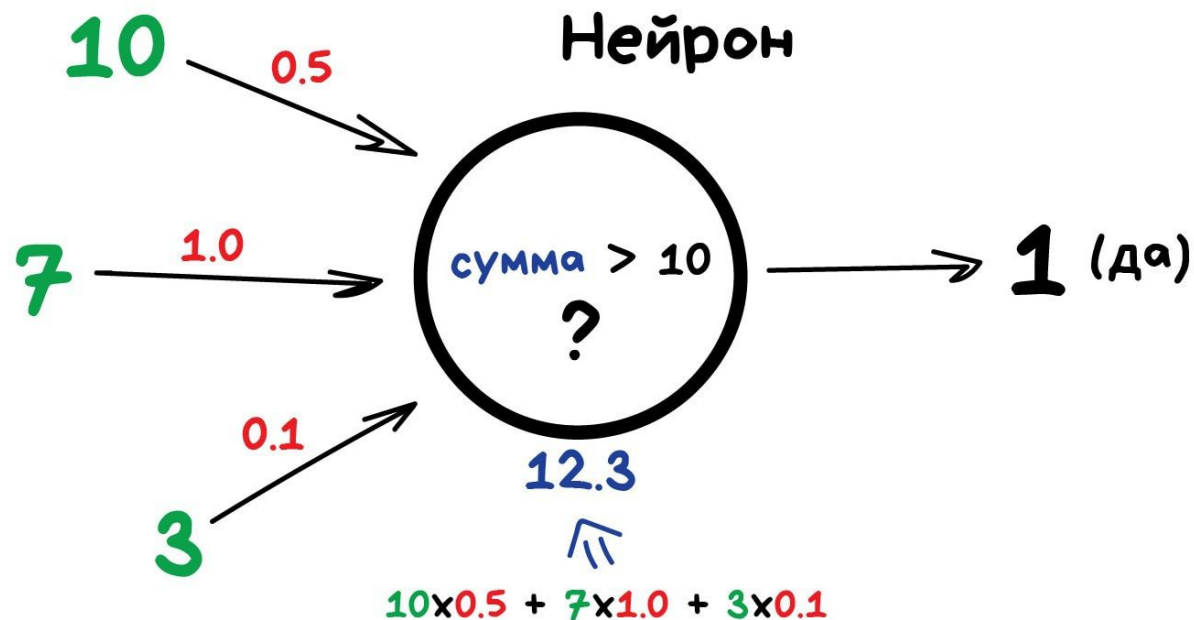


Neural Networks

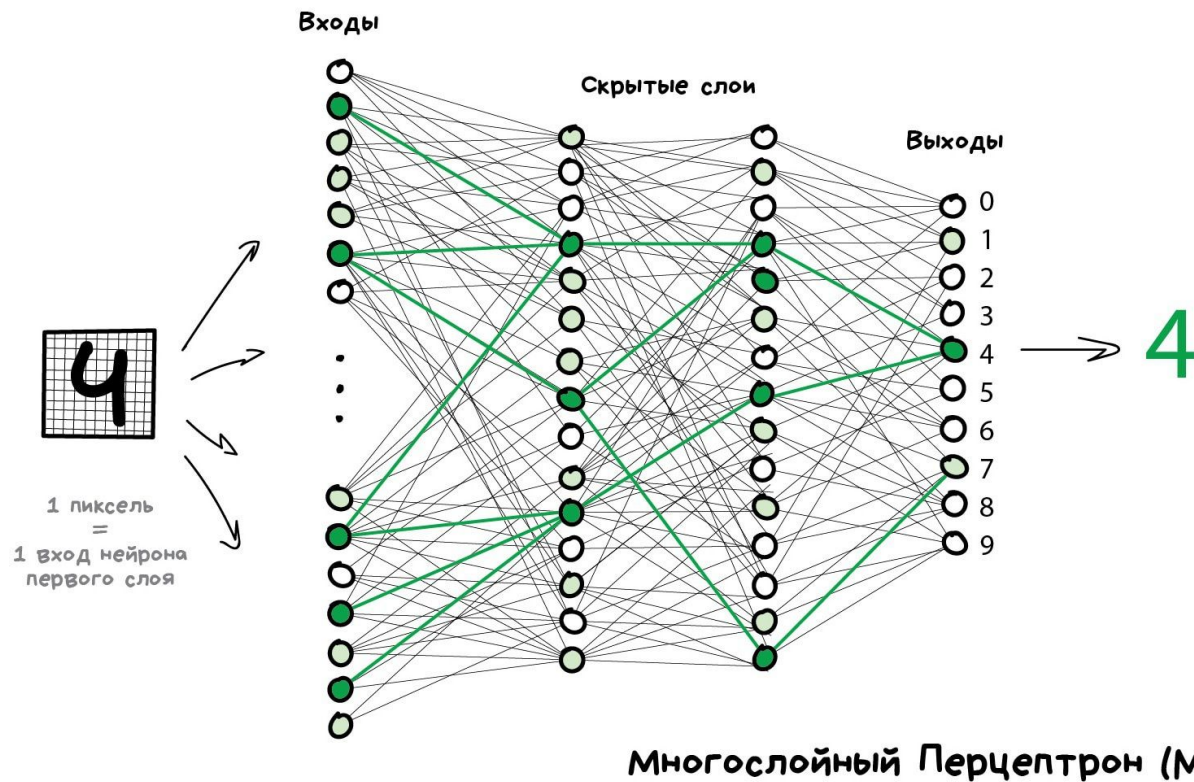
Устройство нейрона человека



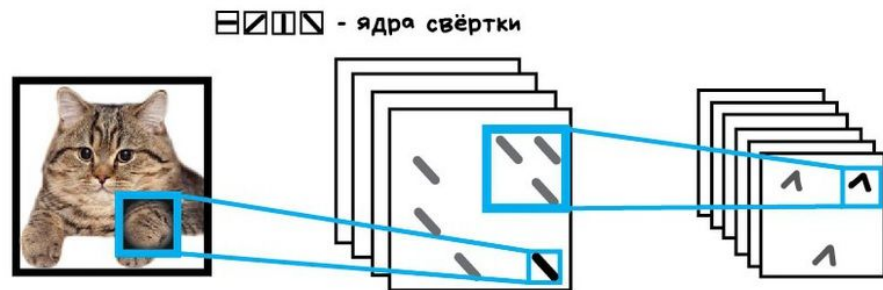
Нейрон в нейронной сети



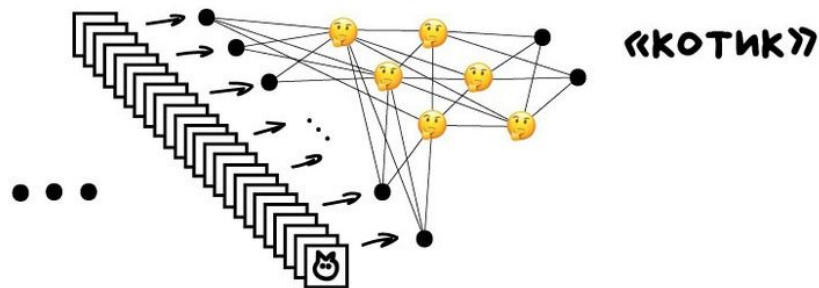
Многослойный перцептрон



Transfer learning



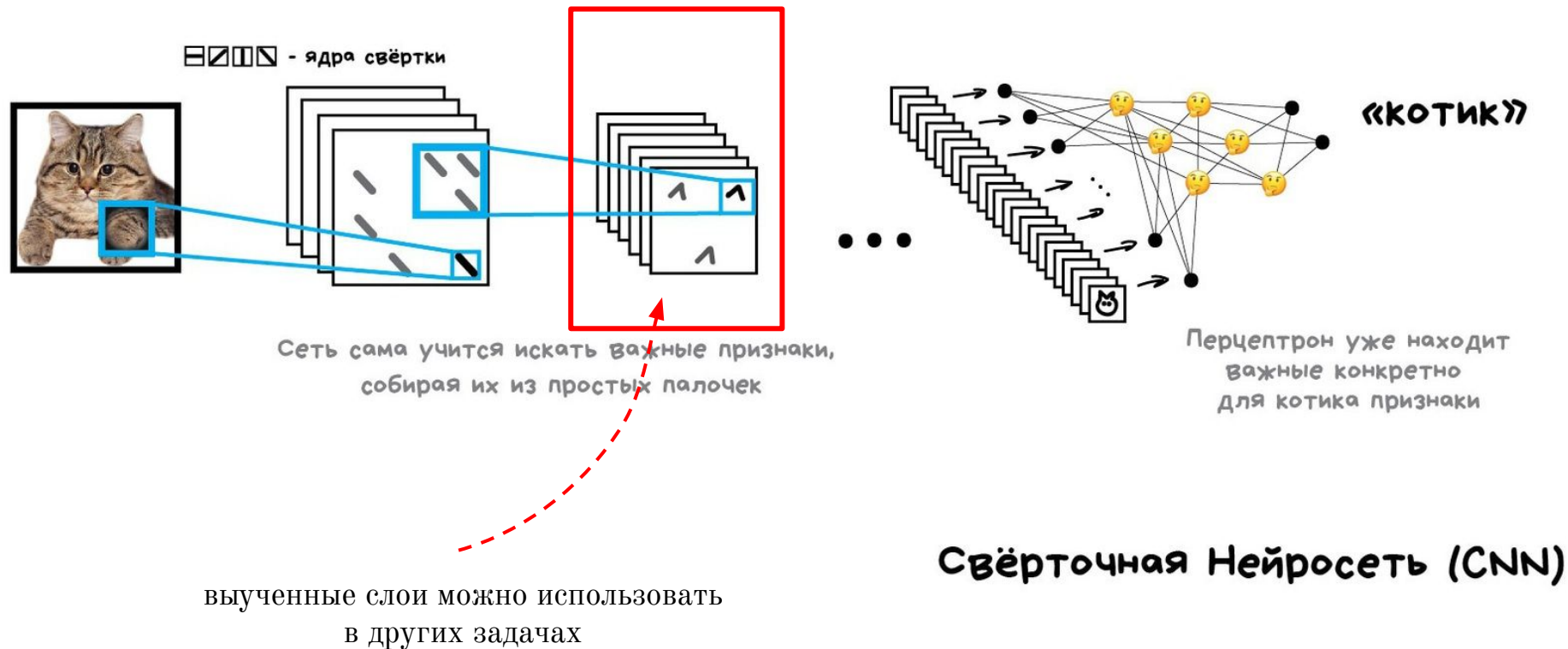
Сеть сама учится искать важные признаки,
собирая их из простых палочек



Перцептрон уже находит
важные конкретно
для котика признаки

Свёрточная Нейросеть (CNN)

Transfer learning



Применение нейронных сетей

Яндекс Переводчик

ТЕКСТ САЙТ КАРТИНКА



АНГЛИЙСКИЙ

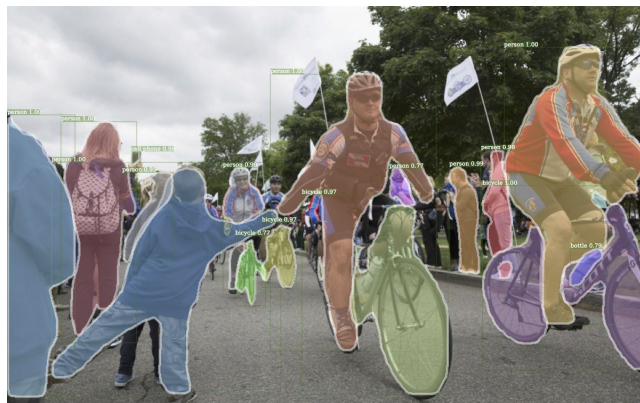


РУССКИЙ



Введите текст или адрес сайта

0 / 10000



Резюме: нейронные сети

- Сложная, но понятная архитектура
- Множество применений, многие из которых ждут своего часа
- Требуют большого количества данных
- Близки к искусственному интеллекту: позволяют осуществлять накопление знаний