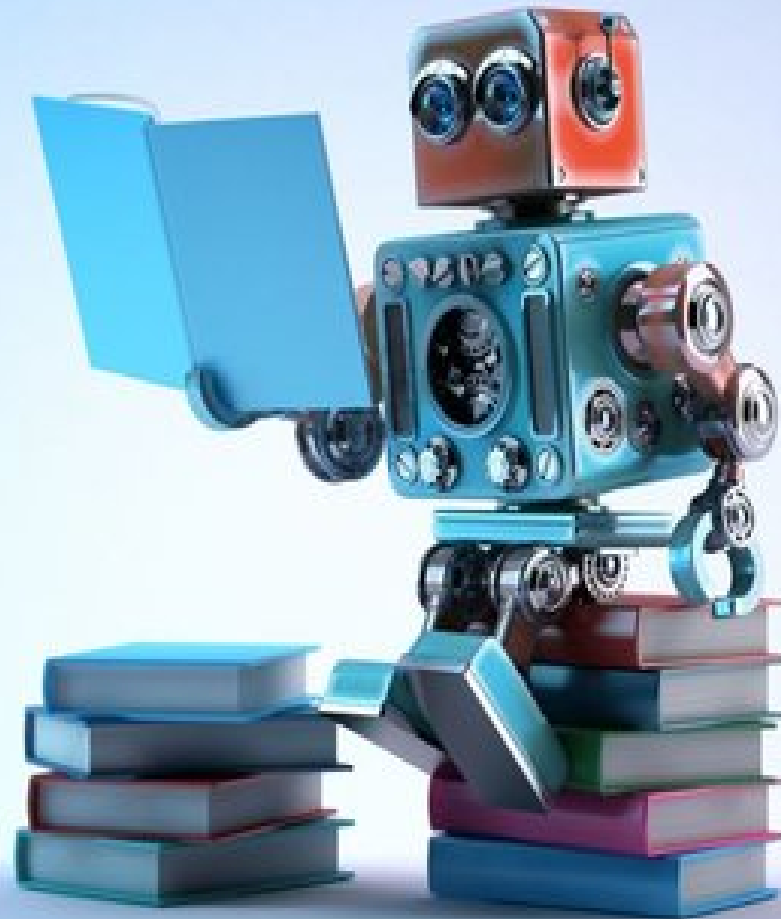


Машинное обучение

Вводная лекция



Калтович Артём
KaltovichArtyom@tut.by
Март 2018

Source: 3 Fundamental Ways Machine Learning Will
Change Business in 2018
<https://www.entrepreneur.com/article/304945>



reCAPTCHA

following finding

and chisels

- 1 августа 2010 - Chad Houck – 31.8%
- 26 мая 2012 – Adam, C-P and Jeffball - 99.1% (аудио версия)
- За несколько часов до презентации Google выпустил обновление, точность упала до 60.95%
- 27 июня 2012 - Claudia Cruz, Fernando Uceda, and Leobardo Reyes - 82%



reCAPTCHA

- Август 2012 – более 90% пользователей находят капчу сложной для ввода
- Май 2016 – [прекращена поддержка](#)
- Август 2017 - Kevin Bock, Daven Patel, George Hughey, Dave Levin - [85.15%](#)
- 31 марта 2018 – выключение



reCAPTCHA

Select all images below
that match this one:



Verify

Шахматы

- Февраль 1996 - Гарри Каспаров 4-2 Deep Blue
- Май 1997 - Гарри Каспаров 2½-3½ Deep Blue
- 2000 - программы Junior и Fritz матчи против Гарри Каспарова и Владимира Крамника – ничья
- Ноябрь-декабрь 2006 года - Владимир Крамник 2-4 Deep Fritz

Го

- Ноябрь 2015 - AlphaGo 4-1 Фань Хуэй
- Март 2016 - AlphaGo 4-1 Ли Седол
- Май 2017 AlphaGo 3-0 Эло Кэ Цзе

В чём же секрет такого успеха?

Машинное обучение

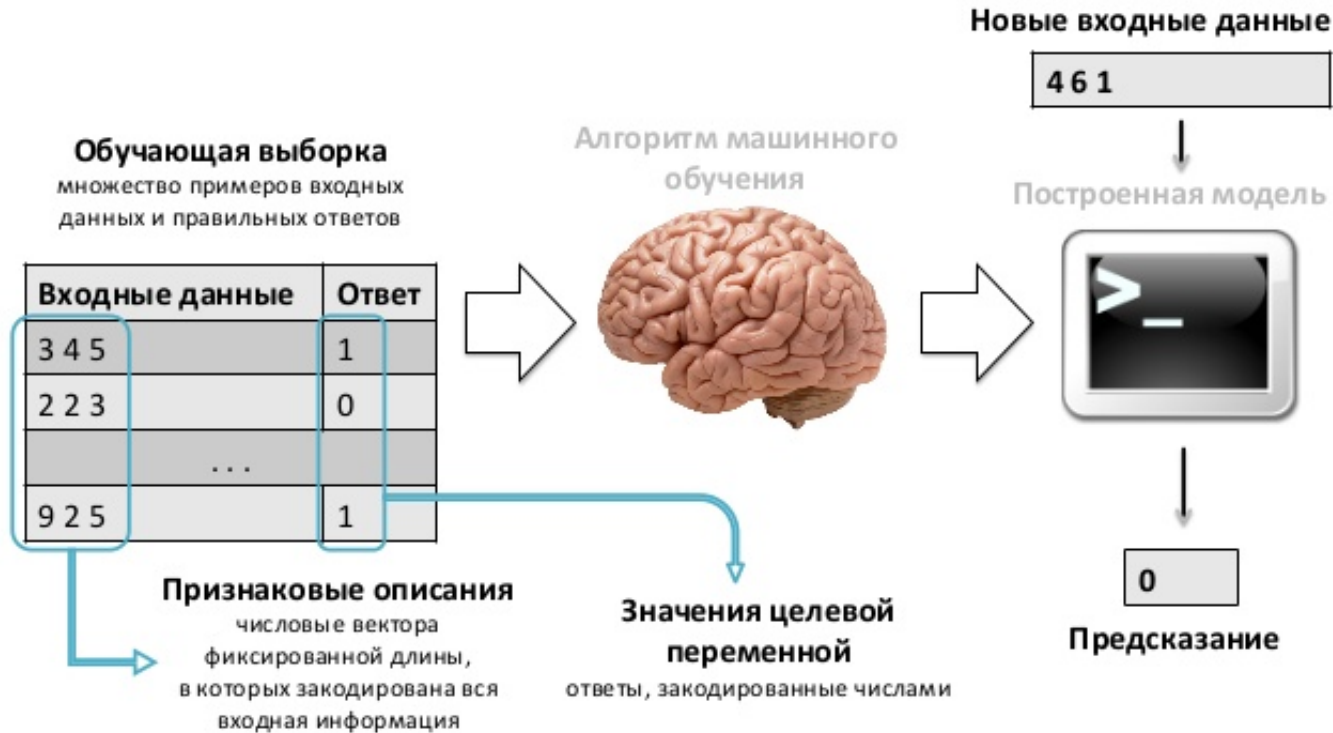
Попробуем сами



Титаник

Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
3	Panula, Mrs. Juha (Maria Emilia Ojala)	female	41.0	0	5	3101295	39.6875	NaN	Southampton
1	Penasco y Castellana, Mr. Victor de Satode 727	male	18.0	1	0	PC 17758	108.9000	C65	Cherbourg
2	Renouf, Mrs. Peter Henry (Lillian Jefferys)	female	30.0	3	0	31027	21.0000	NaN	Southampton
1	Taussig, Mr. Emil	male	52.0	1	1	110413	79.6500	E67	Southampton
3	Peduzzi, Mr. Joseph	male	NaN	0	0	A/5 2817	8.0500	NaN	Southampton

Машинное обучение



Определение

Машинное обучение (англ. machine learning, ML) — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач.

Определение №2

Алгоритм A обучается с эффективностью E над данными D , если при росте мощности $|D|$, E проявляет тенденцию к увеличению.

Примеры применения

- Распознавание номеров автомобилей



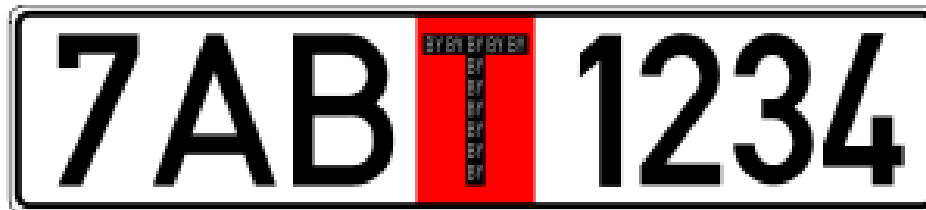
Примеры применения

- Распознавание
номеров
автомобилей



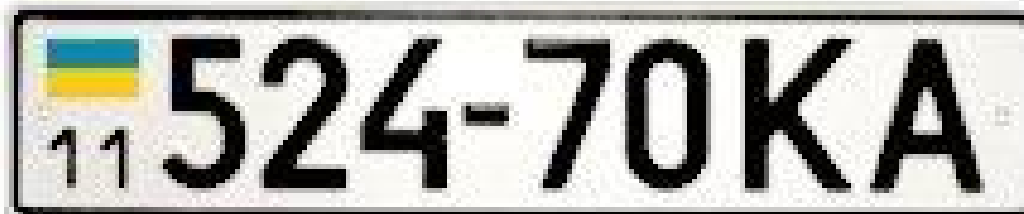
Примеры
применения

- Распознавание
номеров
автомобилей








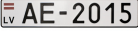




























Примеры применения

- Распознавание
номеров
автомобилей



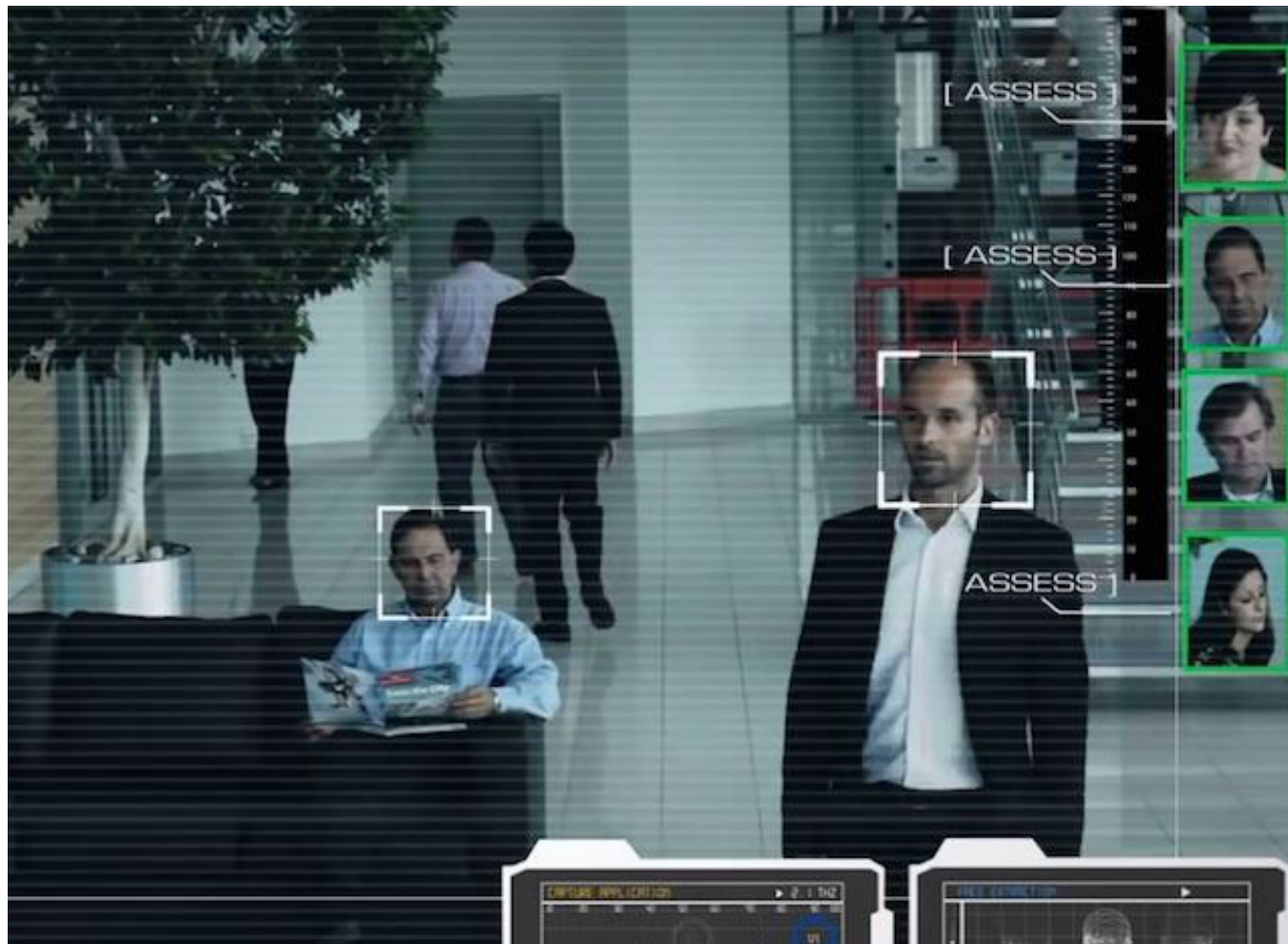
Примеры применения

- Распознавание номеров автомобилей

	Белорусские (BY)		Турция (TR)
	Литовские (LT)		Бельгия (B)
			Дания (DK)
	Латвийские (LV)		Испания (E)
			Португалия (P)
	Эстонские (EST)		Швеция (S)
	Польские (PL)		Финляндия (FIN)
			Швейцария
	Молдавские (MD)		Франция (F)
			Италия (I)
	Приднестровские		Англия (GB)
	Болгарские (BG)		Греция (GR)
			Норвегия (N)
	Немецкие (D)		Голландия (NL) желтая пластина
	Чехия (CZ)		
	Словакия (SK)		Грузия (GEO)
	Венгрия (H)		
	Словения (SLO)		Азербайджан (AZ)
	Сербия (SRB)		
	Хорватия (HR)		Казахстан (KZ)
	Македония (MK)		
	Румыния (RO)		Таджикистан (TJ)
			Туркменистан (TM)
	Австрия (A)		Киргизия
	Албания (AL)		

Примеры применения

- Распознавание лиц

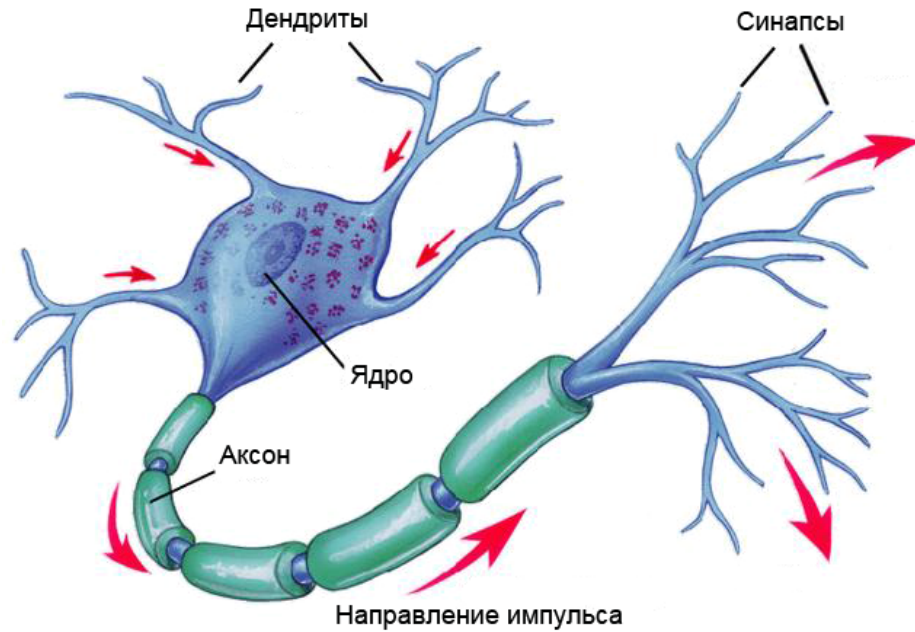


Source: Берлин: супермаркеты Real оборудуют системой распознавания лиц
<https://kriminalnews.de/?p=23123>

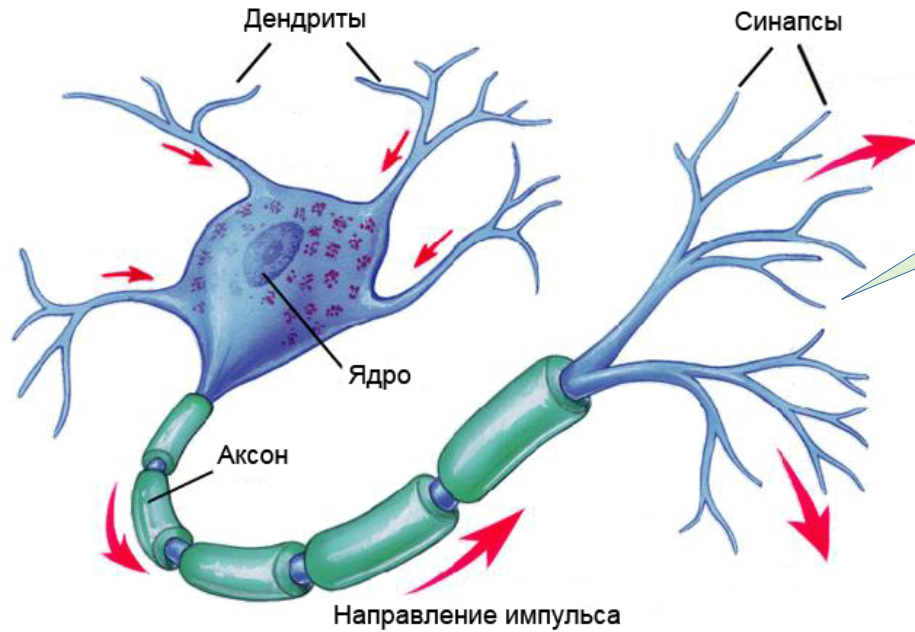
Примеры применения

- Распознавание изображений
- Перевод текстов
- Определение зловредных транзакций
- Фильтрация спама
- ИИ в компьютерных играх
- Высокоскоростная торговля на бирже
- Какие ещё?

Нейронные сети



Нейронные сети



Попробуем сами

Обработка текста

Компьютер может работать только с числами?

Но как превратить слова в числа?

Word2Vec

Но что насчёт контекста?

Что если мы будем анализировать не только само слово, но и его соседей?

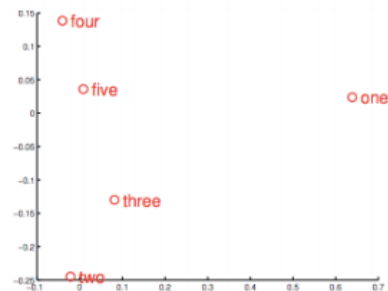
Word2Vec

$\text{vector('Paris')} - \text{vector('France')} + \text{vector('Italy')} \approx \text{vector('Rome')}$
 $\text{vector('king')} - \text{vector('man')} + \text{vector('woman')} \approx \text{vector('queen')}$
 $\text{vector('brother')} - \text{vector('man')} + \text{vector('woman')} \approx \text{vector('sister')}$

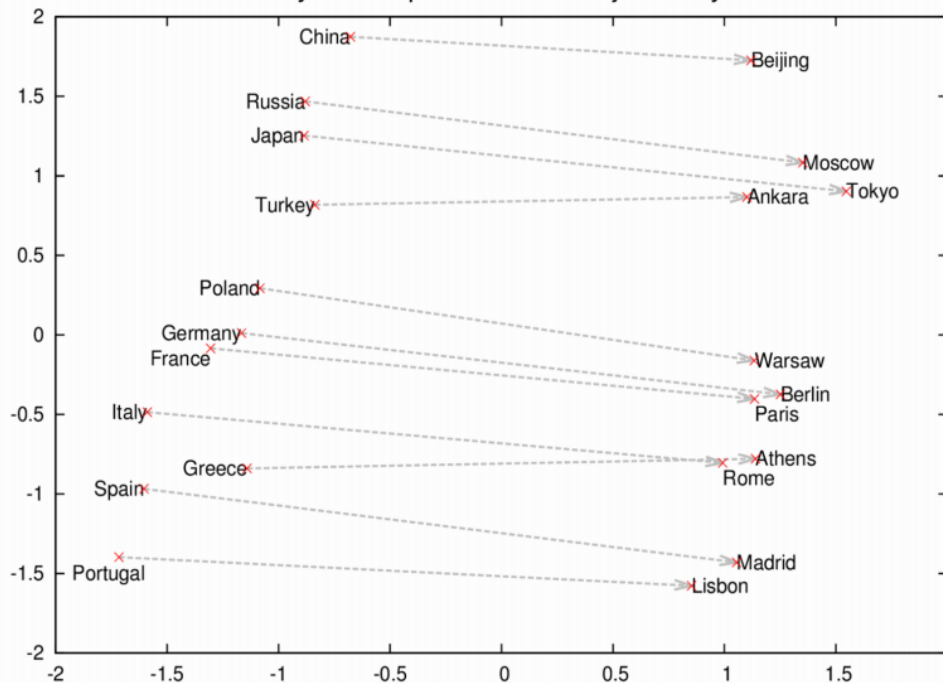
Word2Vec

- Geopolitics: Iraq - Violence = Jordan
- Distinction: Human - Animal = Ethics
- President - Power = Prime Minister
- Library - Books = Hall
- Analogy: Stock Market \approx Thermometer

Word2Vec



Country and Capital Vectors Projected by PCA



Расстояние до Франции :)

spain	0.678515
belgium	0.665923
netherlands	0.652428
italy	0.633130
switzerland	0.622323
luxembourg	0.610033
portugal	0.577154
russia	0.571507
germany	0.563291
catalonia	0.534176

Ресурсы

- <http://scikit-learn.org/> - Документация, содержит много базовой информации и примеров
- <http://playground.tensorflow.org/> - поиграться с нейронной сетью онлайн
- <https://www.kaggle.com/> - сайт для соревнований и тренировок
- habrahabr.ru и geektimes.ru – русскоязычное сообщество
 - <https://geektimes.ru/post/277088/> - статья про нейронные сети
- Конечно же, Википедия – лучше [английская](#).

