

# **Лекция 1: О задачах и методах машинного обучения**

Евгений Борисов

# ML: что взять с собой?

- Мат.анализ
- Алгебра
- Теория вероятностей и Мат.статистика
- Программирование (cuda, python)

# ML: что почитать?

- Andrew Ng - Machine Learning
- Константин Воронцов - Машинное обучение
- Евгений Борисов - <http://mechanoid.kiev.ua>

# ML: с чего все начинается?

извлечение признаков из объекта  
(feature extracting)

формирование пространства признаков

объект -> [FE] -> признаки -> [ML] -> результат

# ML: с чего все начинается?

Классификатор: хорошие и плохие коты

извлекаем признаки

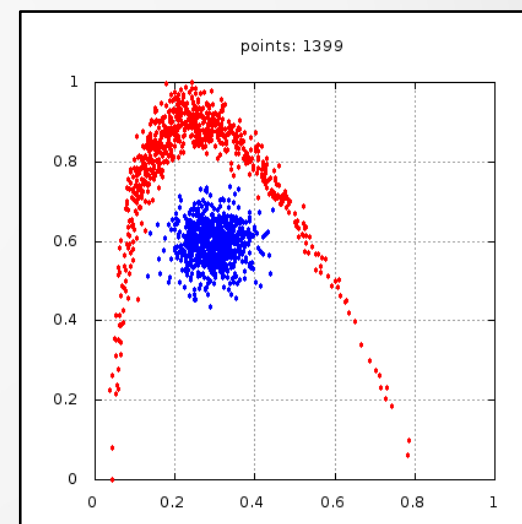
один кот — одна точка



→ [0.14, 12, ..., 345]



→ [78.0, 20, ..., 177]



# ML: обучение

формируем учебный набор

обучаем модель

запускаем модель в работу

# ML: эффект переобучения

хорошо обучается =)

плохо работает :(

# ML: борьба с переобучением

формируем 3 набора  
учебный / контрольный / тестовый

обучаем на учебном  
и проверяем на контрольном

ИТОГОВЫЙ тест на тестовом

запускаем модель в работу



# ML: и что дальше?

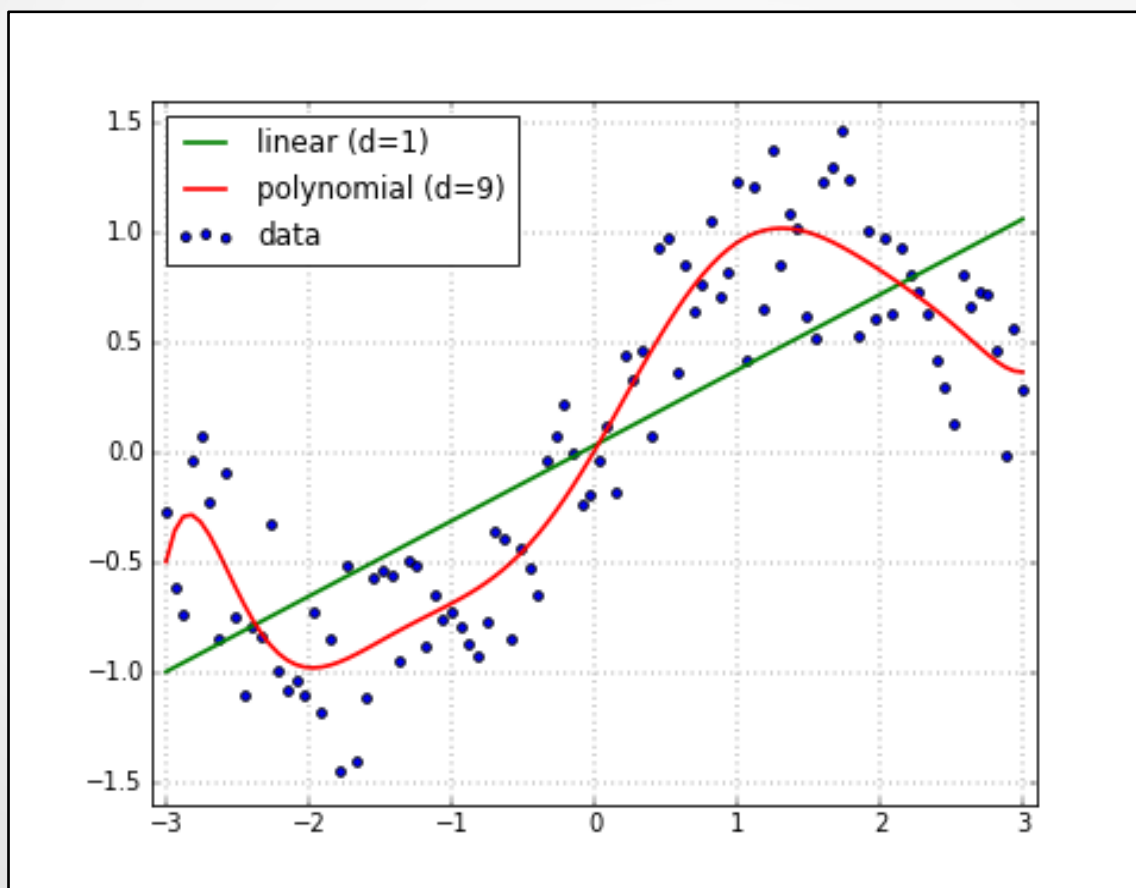
## Задачи:

- Регрессия - восстановление зависимости
- Классификация - разделение на части
- Кластеризация - формирование групп

# ML: регрессия

восстановление зависимости по набору точек

**Оценка недвижимости:** [район, площадь] → цена



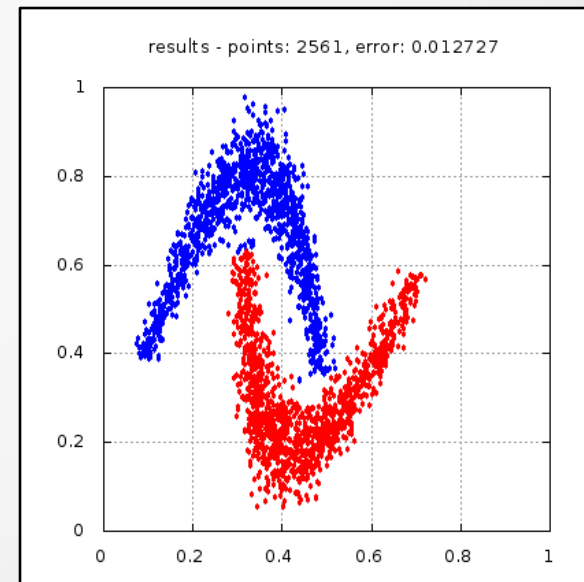
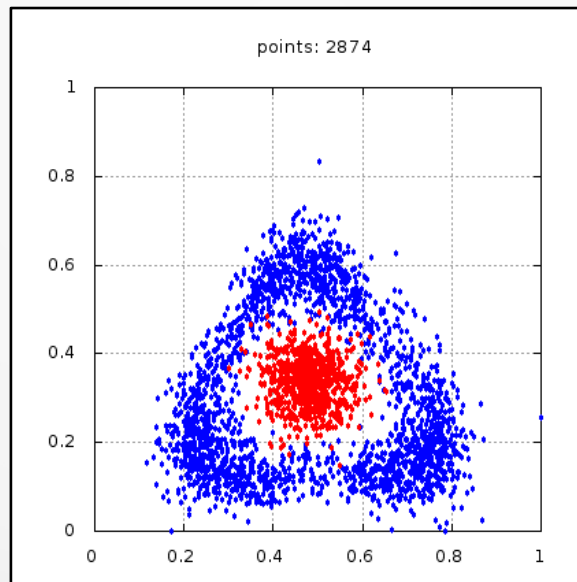
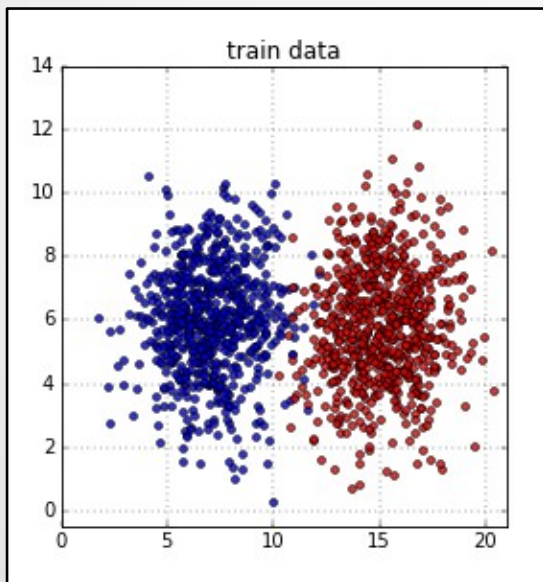
# ML: классификация

разделения объектов на классы

Детектор котов:



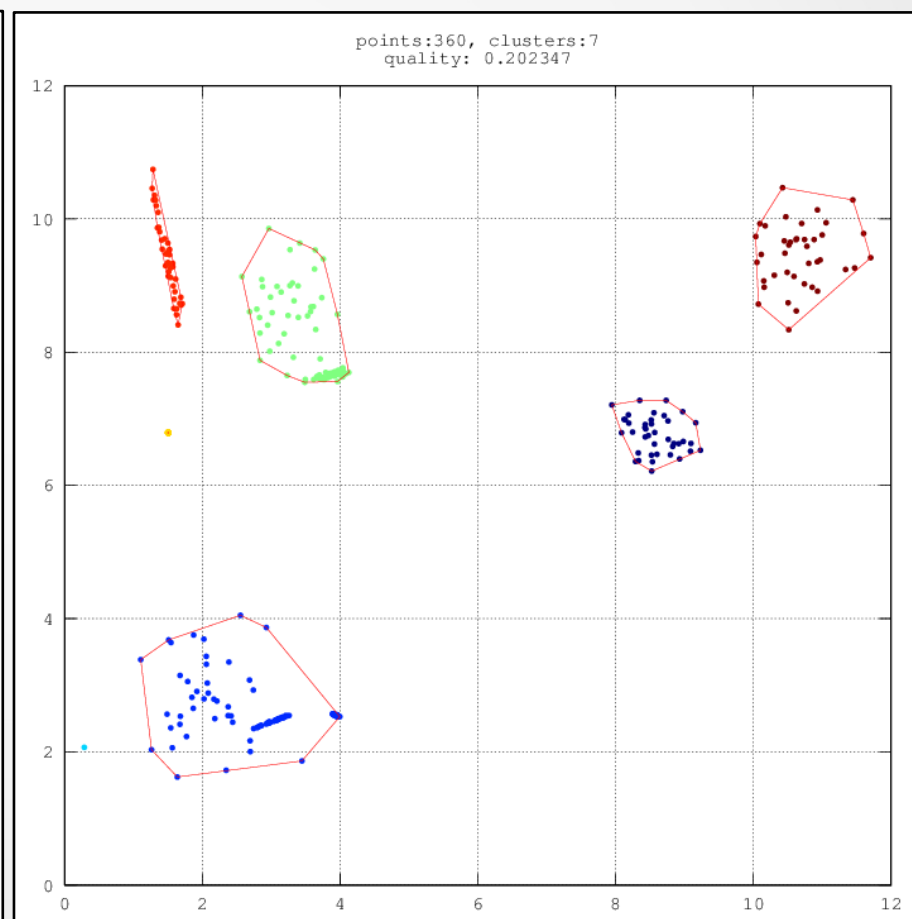
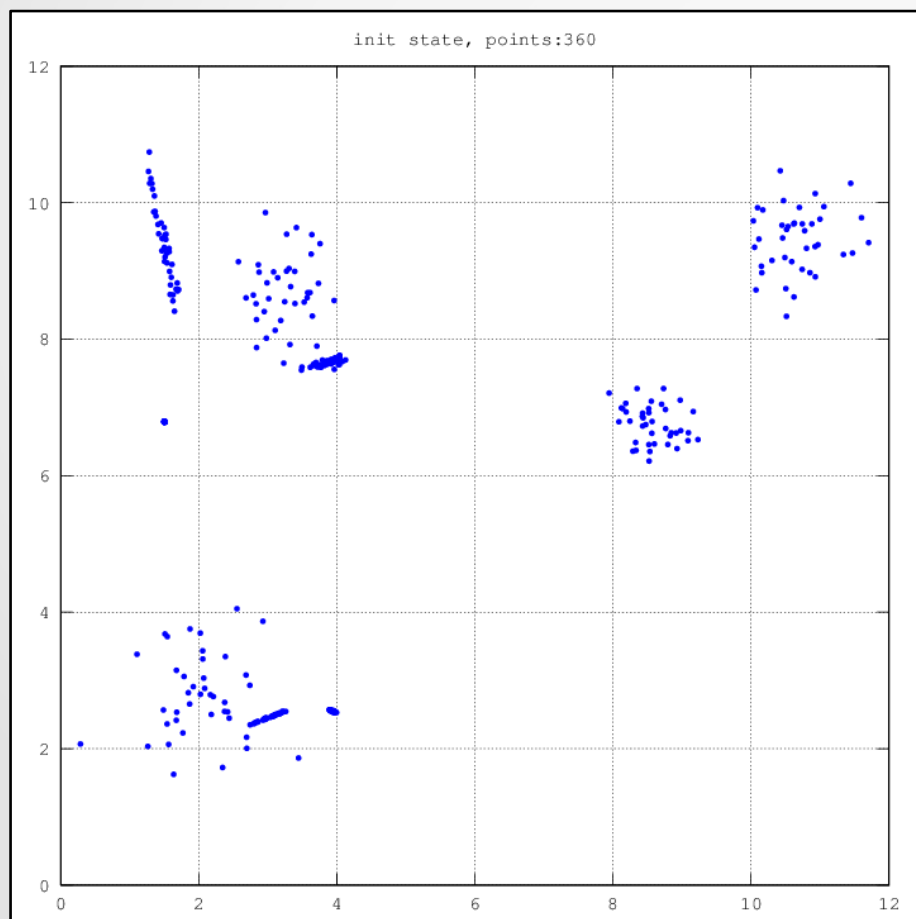
→ вектор-признак → есть/нет



# ML: кластеризация

объединение схожих объектов в группы

**Поиск похожих текстов:** текст → признаки → группа



# ML: и где там обучение?

## Схемы обучения

- с учителем - размеченные данные
- без учителя - данные не размечены
- частичное - данные размечены частично
- с подкреплением - данные заданы неявно

# ML: обучение «с учителем»

**Учебный набор:**  
[ объект, ответ ]

**Задача:** классификатор, регрессия

объект → вектор-признак → результат

**Обучение:** минимизация ошибки

ошибка = результат — правильный ответ

**Критерий остановки:**

достигнут порог значения ошибки,  
и/или порог количества циклов

# ML: обучение «без учителя»

**Учебный набор:**  
[ объект ]

**Задача:** кластеризация

объект → вектор-признак → результат

**Обучение:** изменение параметров

**Критерий остановки:**

состояние не изменяется,  
и/или порог количества циклов

# ML: частичное обучение

**Учебный набор:** частично размечен

[ объект, ответ ] + [ объект ]

**Задача:** классификатор, регрессия, кластеризатор

объект → вектор-признак → результат

**Обучение:** кластеризация + классификатор



# ML: обучение с подкреплением

## Учебный набор:

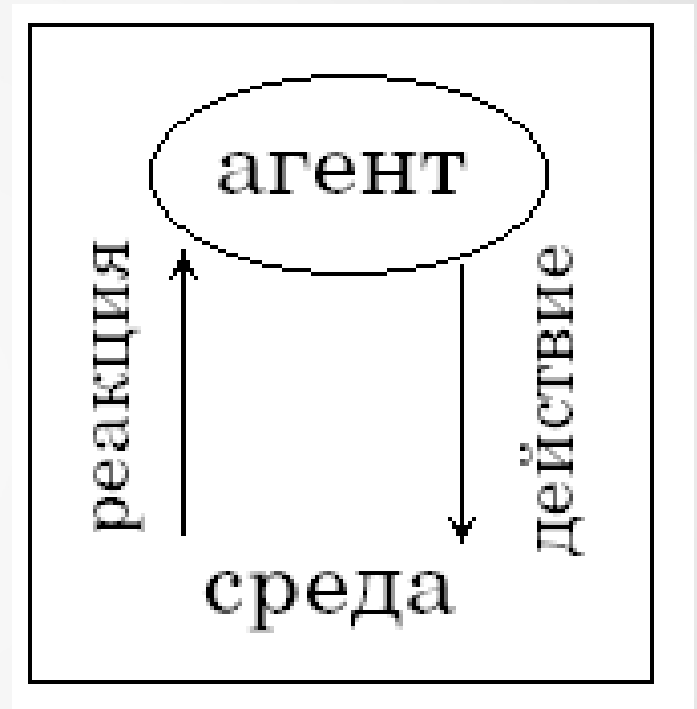
В явном виде отсутствует

## Задача: управление

- предсказываем реакцию среды
- выбираем оптимальное действие

## Обучение:

аппроксимация ф-ции оценки действия



## ML: и куда дальше?

- Статистические: *naïveBayes*, EM
- Логические: *decision tree*
- Метрические: *k-neighbors*, *k-means*
- Линейные: *MLP*
- Композиции: *AdaBoost*
- *Deep Learning*

# ML: зачем это?

## Области применения

NLP

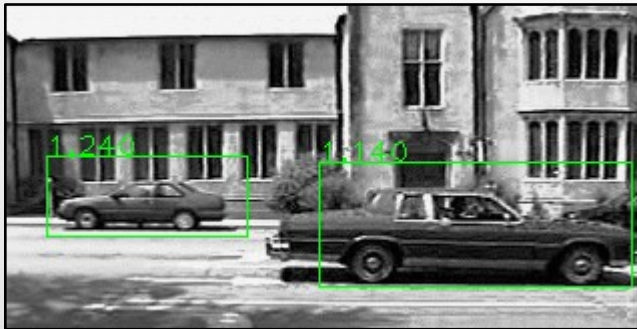
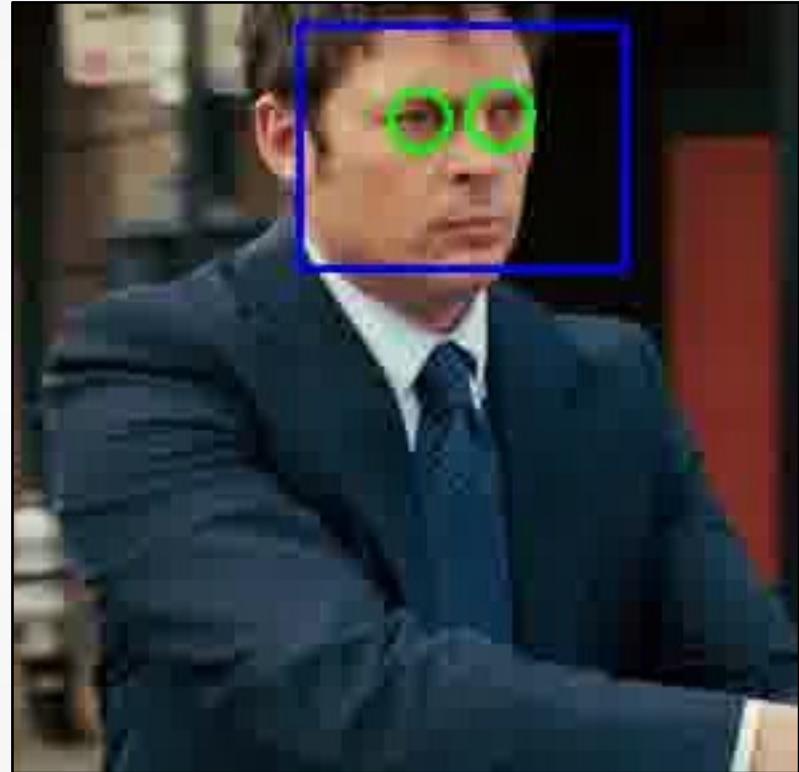
Speech recognition

CV

Data Analysis

...

# ML: Computer vision (CV)



# ML: Natural Language Processing (NLP)

## Поиск похожих текстов

Около 18 тысяч человек покинули подконтрольные боевикам районы Алеппо. За минувшие сутки из подконтрольных боевикам районов сирийского города Алеппо было выведено около 17,971 тысячи жителей, в их числе 7,542 тысячи детей. Об этом в субботу, 10 декабря, сообщает ТАСС со ссылкой на российский Центр примирения враждующих сторон в Арабской Республике.

Битва за Алеппо: повстанцы просят дать им вывезти раненых  
Сирийские повстанцы просят о пятидневном перемирии, чтобы эвакуировать раненых из районов в восточной части Алеппо, после того как они вывели все свои отряды из исторического центра — Старого города.

# ML: технические средства

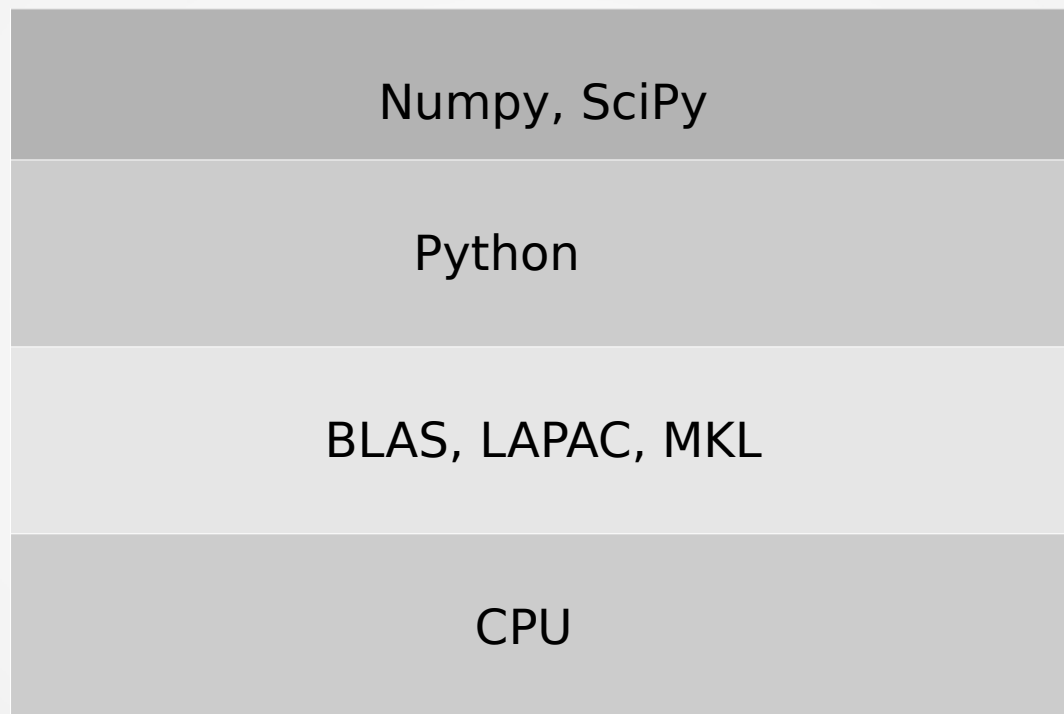
прикладные программные средства

вычислительные библиотеки

программный интерфейс с аппаратурой

аппаратные вычислительные средства

# ML: технические средства



# ML: технические средства

Keras

Theano, TensorFlow

cuBLAS, cuDNN, CNMeM, NCCL

CUDA

NVIDIA GPU



# ML: технические средства

Python

Keras

OpenCV

Octave

Theano

scikit-image

Jupyter

TensorFlow

Numpy

GPU

Matplotlib

CUDA

Scikit-Learn

OpenCL

NLTK

ML: ну, наконец-то! последний слайд...

**Вопросы ?**