



О методах машинного обучения

Евгений Борисов

О методах ML

Как это работает ?

формируем учебный набор

обучаем модель

запускаем модель в работу

О методах ML

...а чтобы сам учился ?

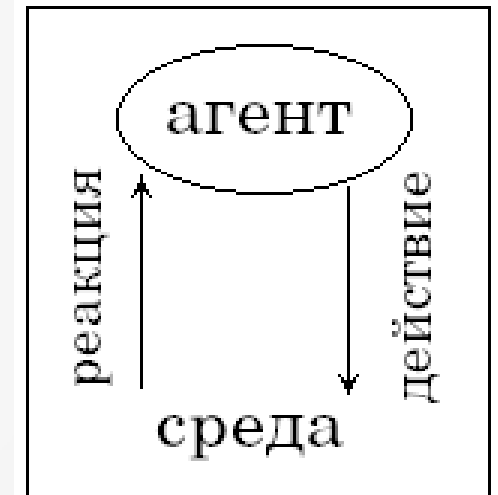
обучение с подкреплением

учебного набора в явном виде нет

собираем историю действий и последствий

пытаемся предсказывать реакцию среды

выбираем оптимальное действие



О методах ML

извлечение признаков из объекта
(feature extracting)

формирование пространства признаков

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

О методах ML

датасет - размеченная матрица признаков

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} & y_1 \\ x_{21} & x_{22} & \dots & x_{2n} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} & y_m \end{bmatrix}$$

x - вектор-признак

y - метка класса

n - размер пространства признаков

m - количество примеров

О методах ML

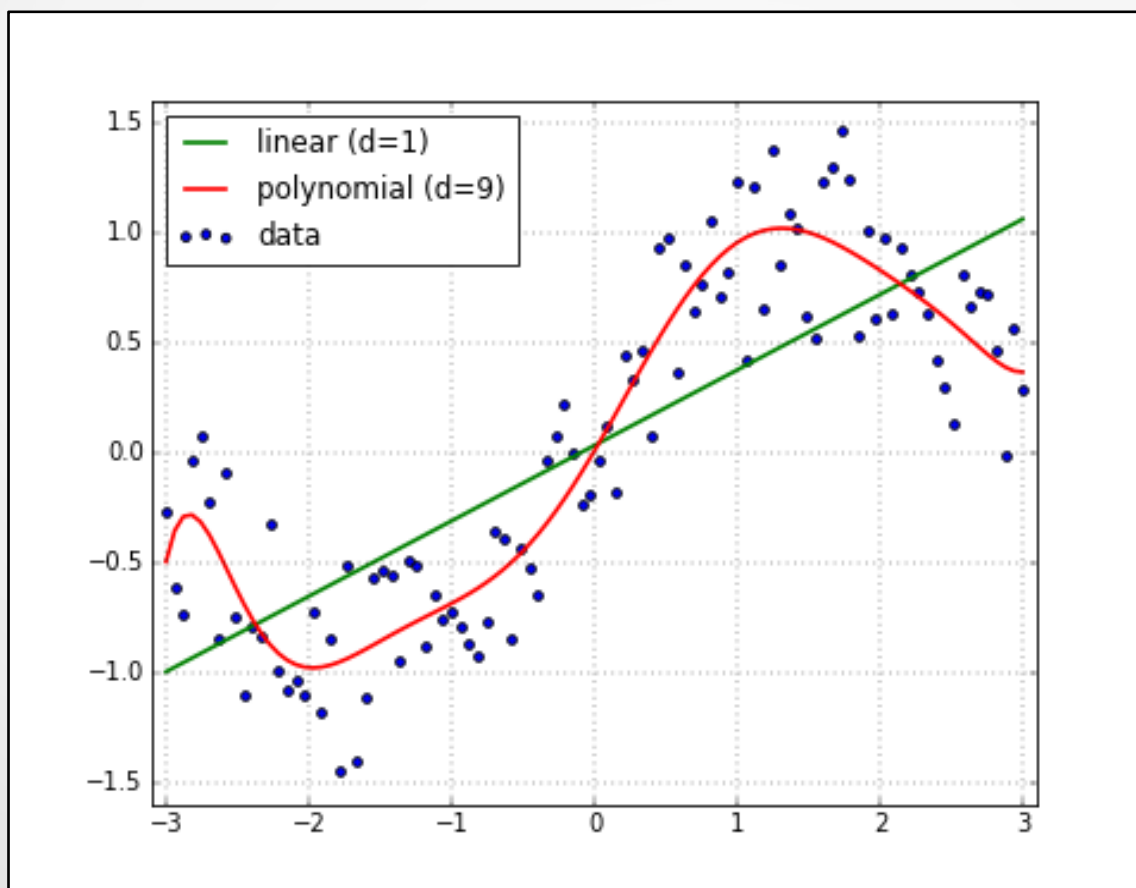
Основные типы задач ML

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

ML: регрессия

восстановление зависимости по набору точек

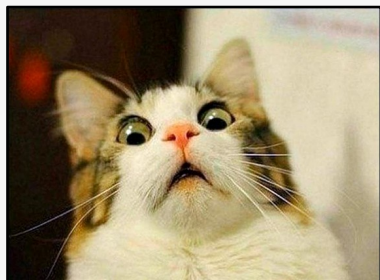
Оценка недвижимости: [район, площадь] → цена



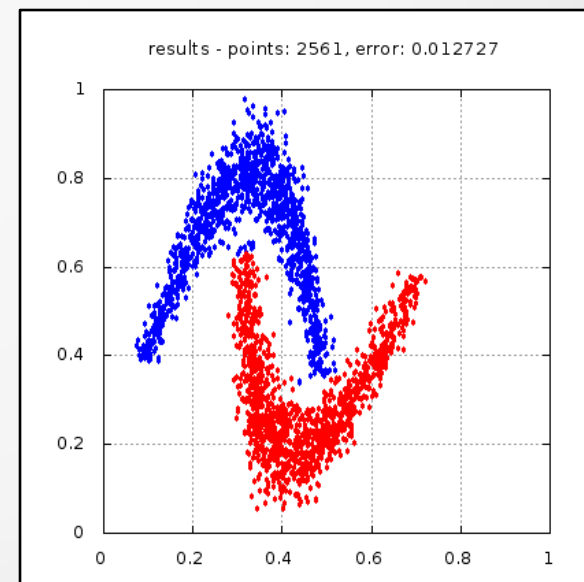
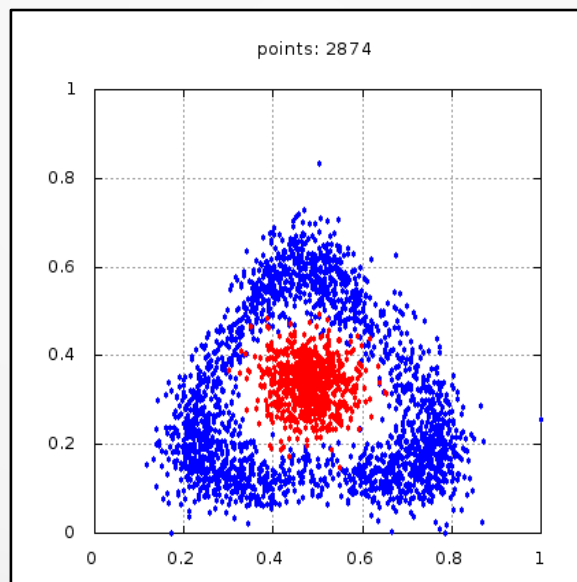
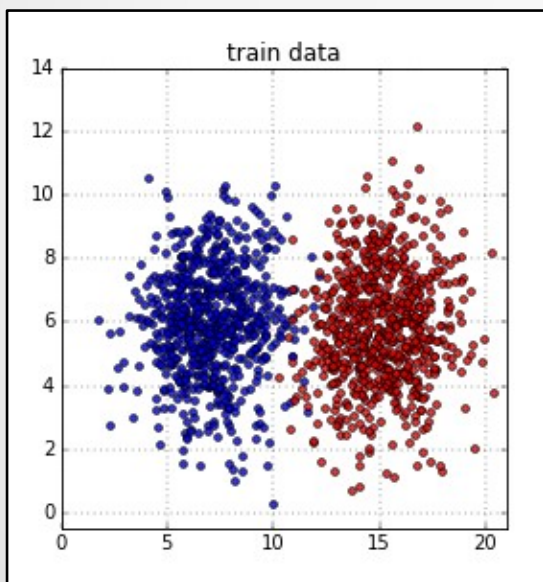
ML: классификация

разделения объектов на классы

Детектор котов:



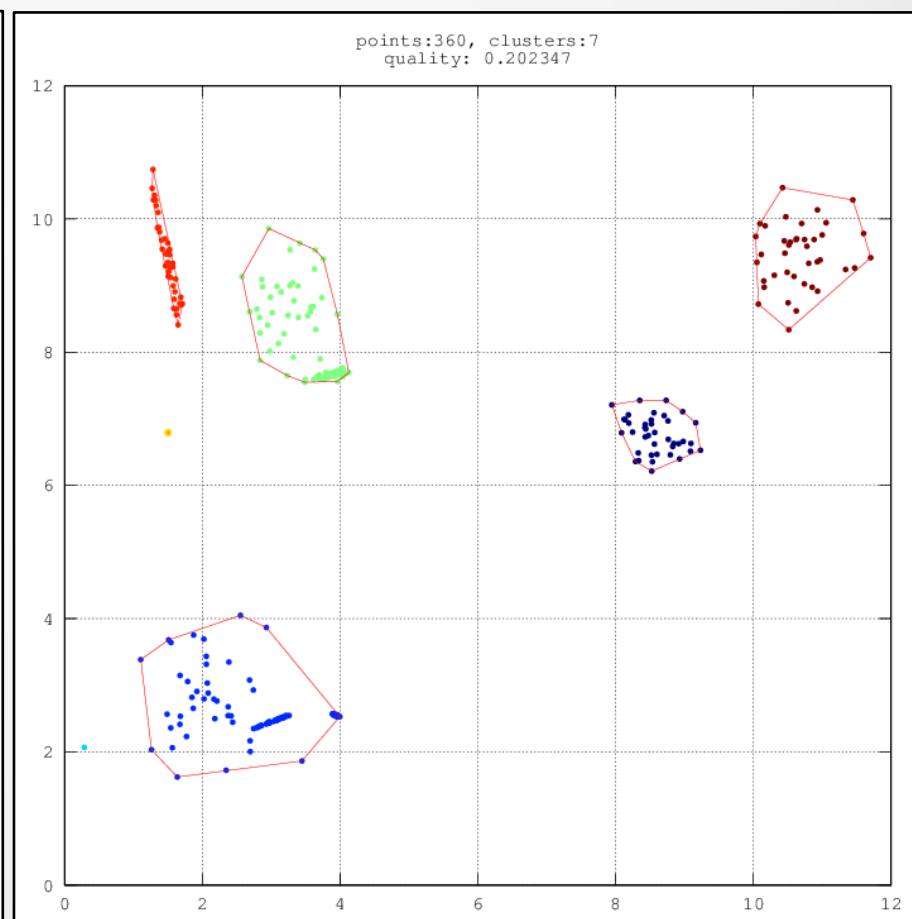
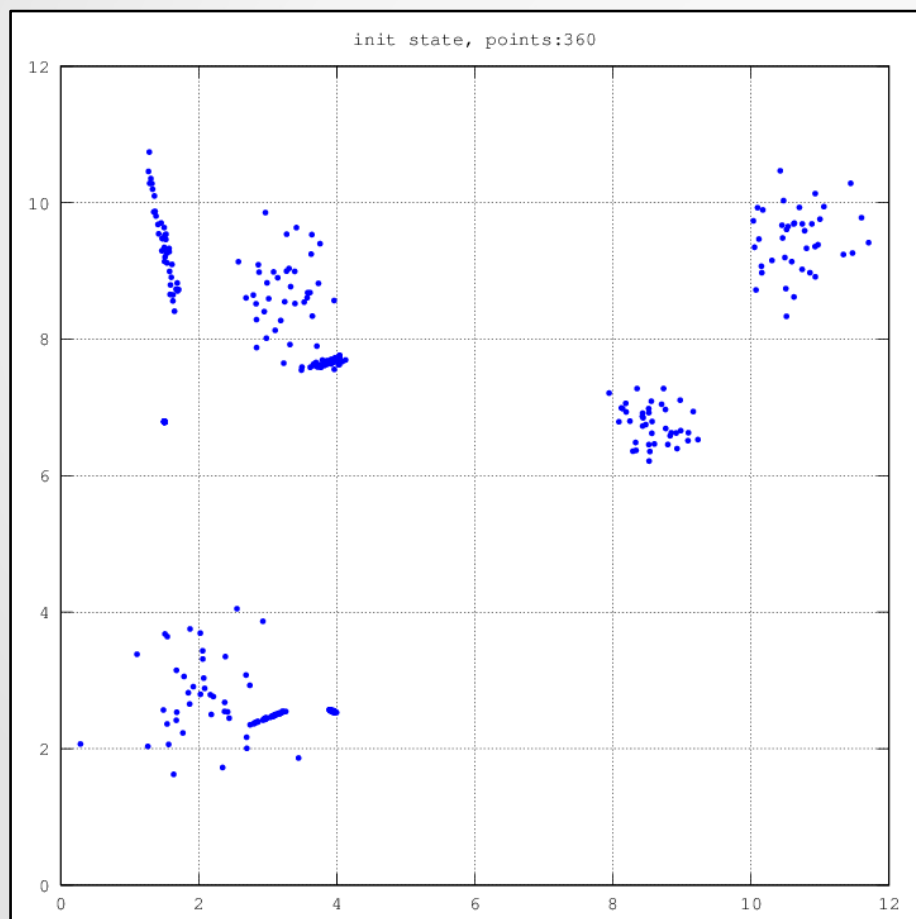
→ вектор-признак → есть/нет



ML: кластеризация

объединение схожих объектов в группы

Поиск похожих текстов: текст → признаки → группа



О методах ML

способы организации данных

"с учителем" supervised

- размеченные данные

"без учителя" unsupervised

- не размеченные данные

"частичное обучение" semi-supervised

- частично размеченные данные

"с подкреплением" reinforcement

- датасет в явном виде отсутствует

подходы и модели

Метрические: k-Neighbors

Статистические: Naive Bayes

Логические: Decision Tree

Линейные: SVM, MLP

Композиции: AdaBoost

Коннекционизм и нейросети

Deep Learning

ML: что почитать?

- Константин Воронцов - Машинное обучение
- Радослав Нейчев - Машинное обучение, ФПМИ, 2020
- Andrew Ng - Machine Learning
- Евгений Борисов - <http://mechanoid.su>
- http://github.com/mechanoid5/ml_lectorium

О работе в Data Science



Вопросы ?