Лекция 15:
Обзор рассмотренных моделей и подведение итогов.

Евгений Борисов

ML обзор

Список рассмотренных тем

общая схема применения методов ML

основные типы задач

извлечение признаков и формирование датасета

оценка качества классификаторов и выбор модели

статистические методы

метрические методы

линейные методы

логические методы

композиции классификаторов

ML обзор

Общая схема применения методов ML

определяем задачу

изучаем предметную область

формализуем задачу

извлекаем признаки из объекта

подбираем преобразования признаков

отбираем хорошие признаки, собираем учебный набор

удаляем выбросы

обучаем модель

тестируем модель

запускаем модель в работу

ML обзор

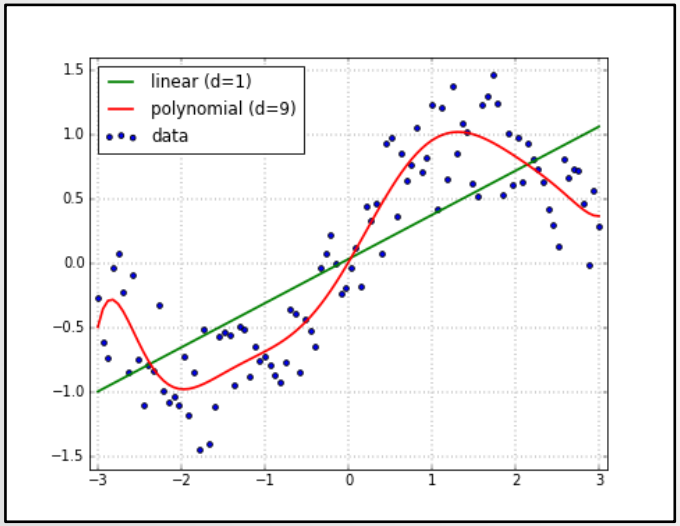
Формализация задачи

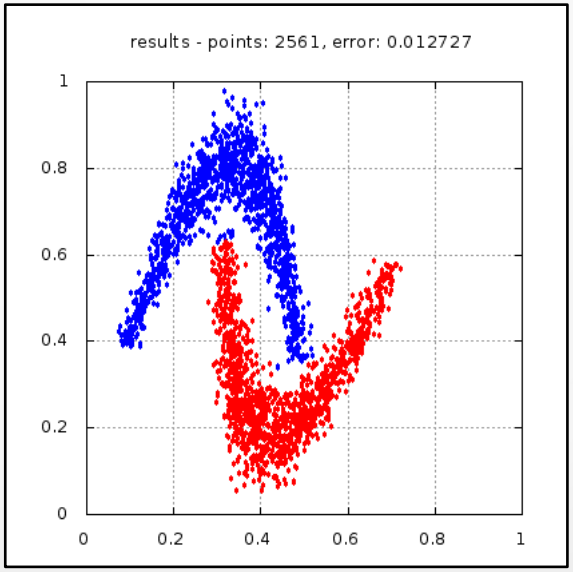
основные типы задач

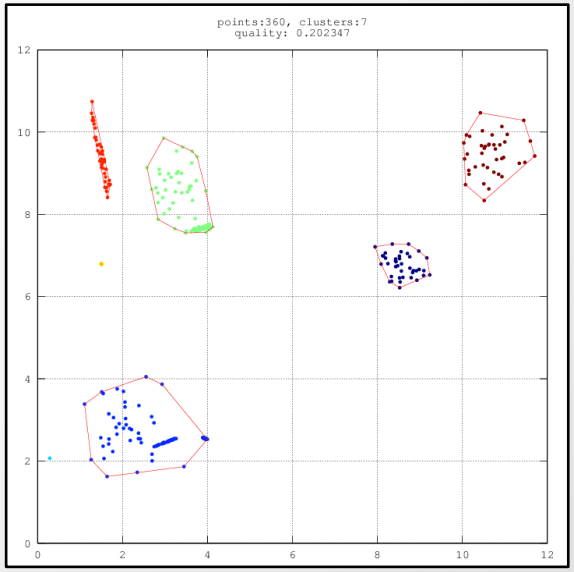
регрессия - восстановление зависимости

классификация - разделение на части

кластеризация - формирование групп







ML обзор

Отбор признаков и выбор модели

извлечение признаков

очистка датасета

трансформации признаков

отбор признаков

ML обзор

Оценка качества классификаторов

погрешность (accuracy)

матрица ошибок (confusion matrix)

точность (precision)

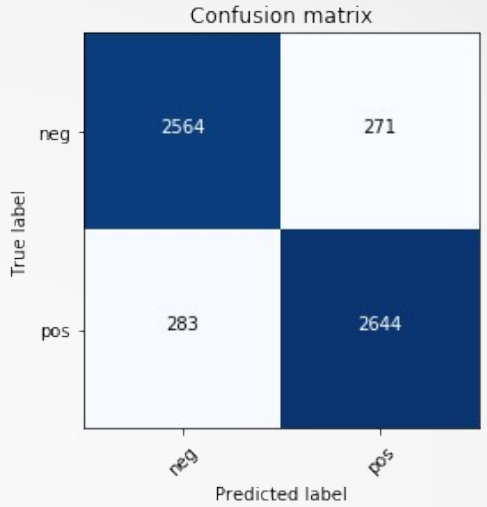
полнота (recall)

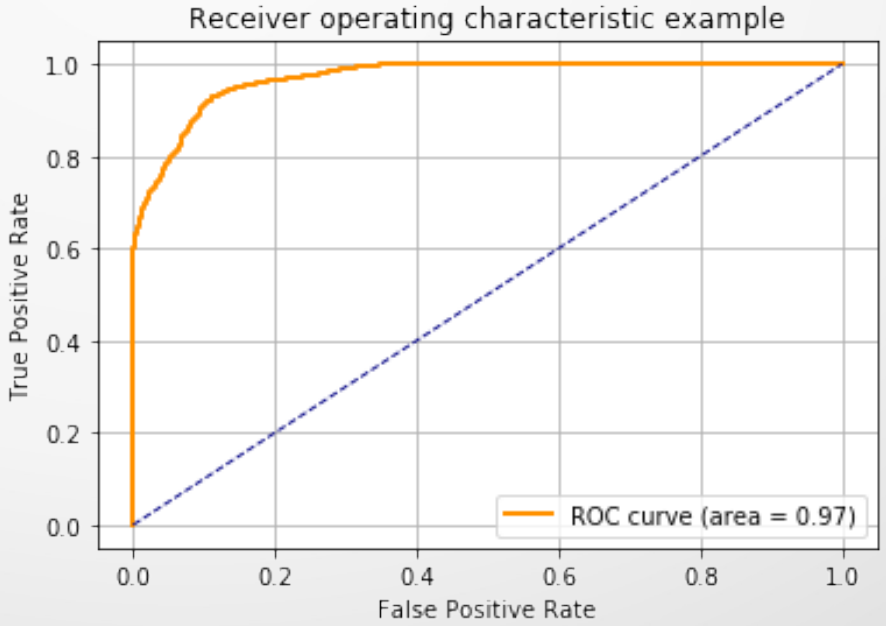
F-мера

ROC/AUC

кросс-валидация

методы выбора моделей





ML обзор

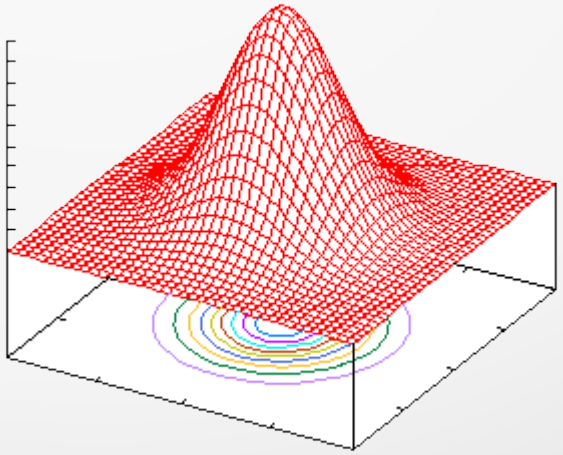
Статистические методы

байесовский классификатор $$a(x)=\underset{y\in Y}{argmax}\,\lambda_y P(y)\,p(x|y)$$

методы восстановления плотности распределения

метод парзеновского окна

EM-алгоритм



ML обзор

Метрические методы

типы кластеров

kNN

k-means

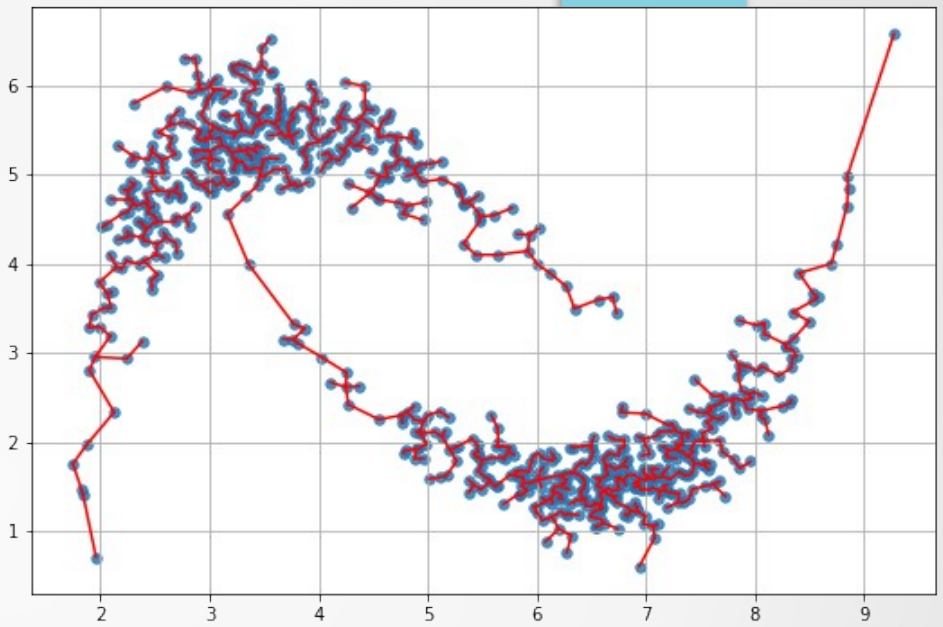
КНП

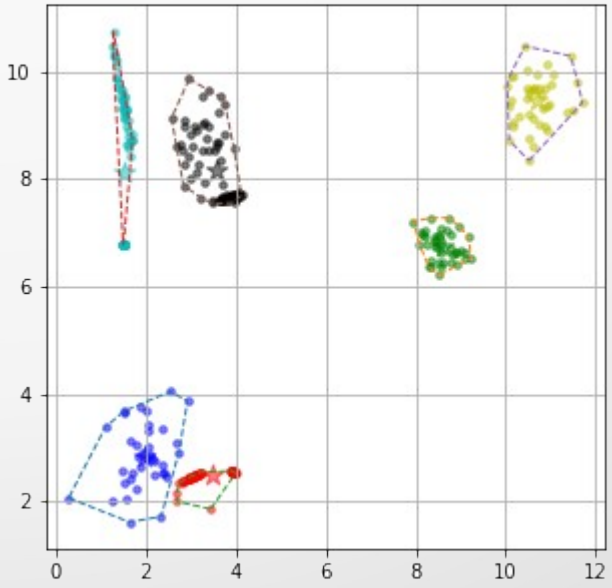
FOREL

DB-SCAN

иерархическая кластеризация

профиль компактности

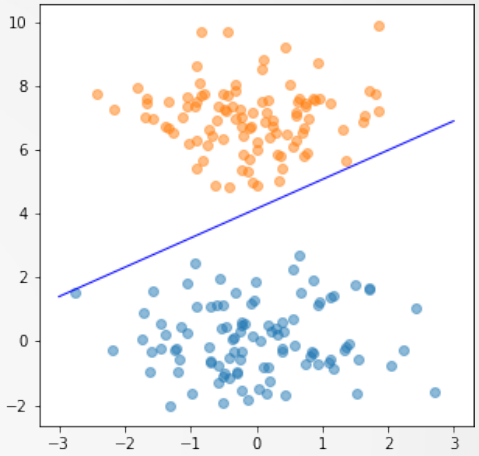




ML обзор

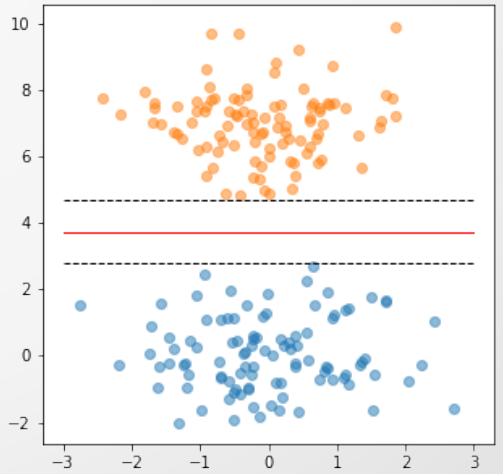
линейные методы

линейный классификатор $$a(x,w)=sign\left(\sum_{i=1}^{n} x_i \cdot w_i - w_0\right)$$

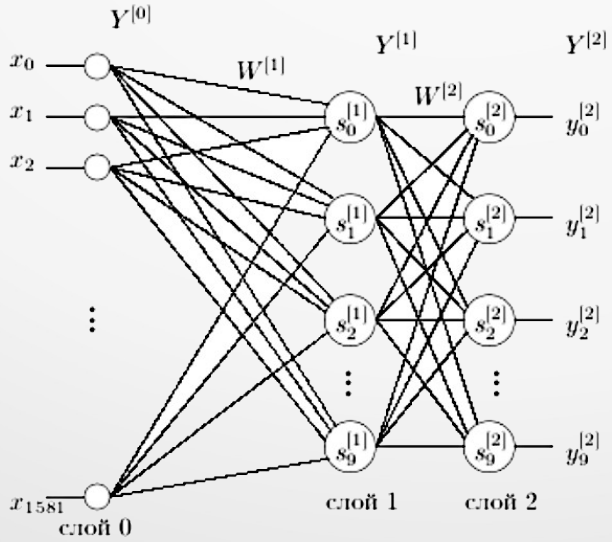


отступ от разделяющей поверхности

SVM $$a(x)=sign\left(\sum_{i} \lambda_i y_i K(x_i,x) - w_0\right)$$



MLP и back-prop



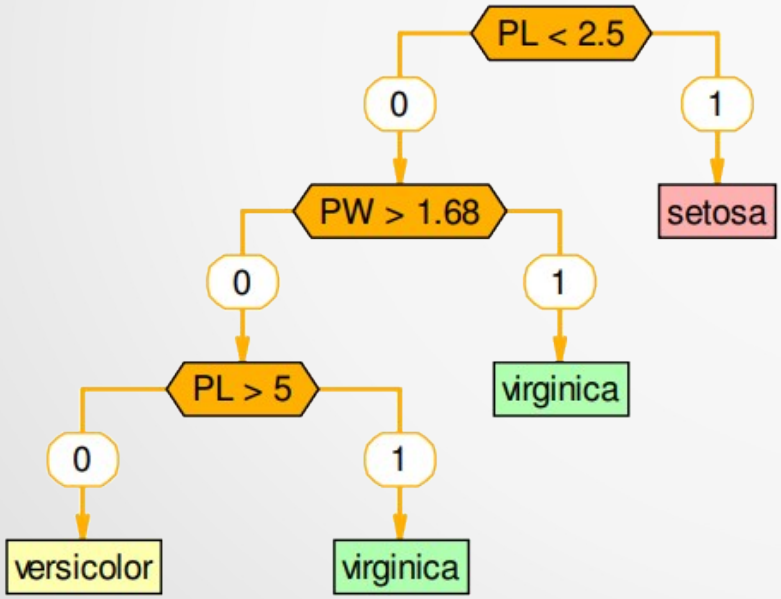
ML обзор

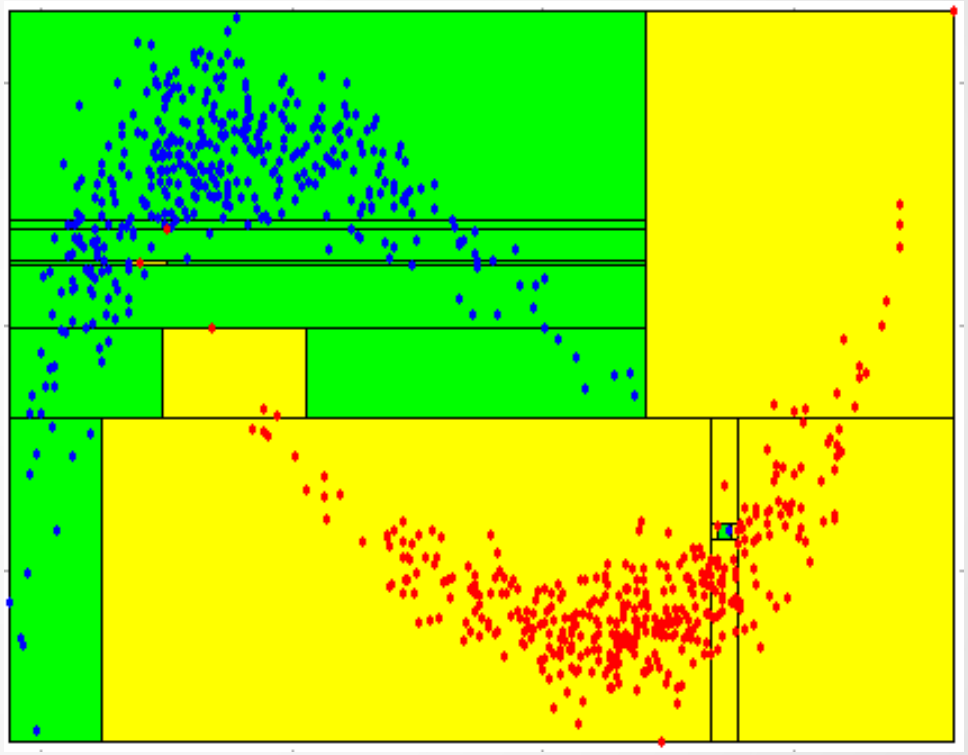
логические методы

информативность iGain, Gini

решающие деревья

оценка важности признаков





ML обзор

Композиции классификаторов

AdaBoost $$a(x) = sign\left(\sum_{i=1}^{T} a_t \cdot b_t(x)\right)$$

bagging - обучение по случайным подвыборкам

rsm - обучение на случайном подмножестве признаков

ML обзор

Литература

Andrew Ng Machine Learning. - Coursera / Stanford University

К.В. Воронцов Машинное обучение - ШАД Яндекс 2014

http://www.machinelearning.ru

Sebastian Raschka Python Machine Learning - Packt Publishing Ltd, 2015

git clone *https://github.com/mechanoid5/ml_lectorium.git*

Евгений Борисов Методы машинного обучения
http://mechanoid.kiev.ua

ML обзор

Вопросы ?