

Дисклеймер

Данная презентация сформирована на основе материалов кафедры ММП факультета ВМК, где читается аналогичный курс

Что такое машинное обучение

- Машинное обучение нацелено на поиск закономерностей в данных
- Данные – совокупность объектов (пользователи, изображения, сигналы, цепочки ДНК, кредитные истории и т.д.)
- Каждый объект представлен набором наблюдаемых переменных (характеристик) и скрытых переменных
- Известна обучающая выборка – совокупность объектов с известными характеристиками и значениями скрытых переменных
- Цель – найти способ предсказать скрытые переменные для нового объекта с известными наблюдаемыми характеристиками

Примеры задач машинного обучения

Алгоритм прогнозирования визитов покупателей супермаркетов и сумм их покупок



Дано: статистика посещений клиентами магазинов сети и суммы их покупок

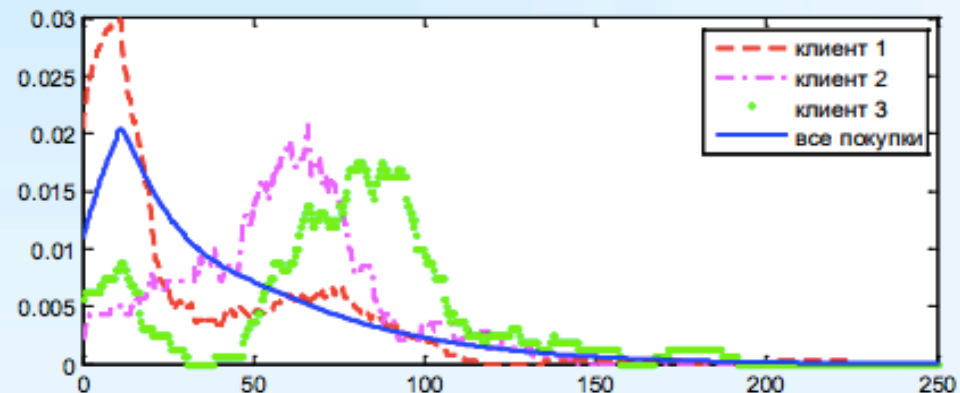
Найти: для каждого клиента — день следующего визита и сумму покупок

Метод: оценка вероятностей визитов и плотности распределения сумм покупок, максимизация правдоподобия

#	Team Name	\$10,000 • 279 teams	Score ?	Entries
1	D'yakonov Alexander (MSU, Moscow, Russia) *		18.83	68
2	NSchneider *		18.67	20
3	Ben Hamner *		18.57	19
4	William Cukierski		18.44	75

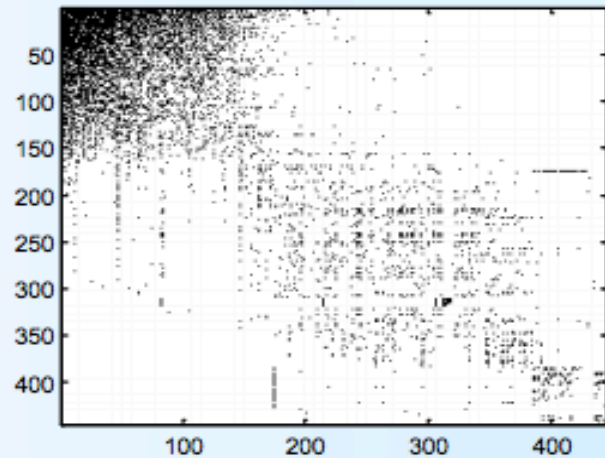
<http://www.kaggle.com/c/dunnhumbychallenge/>

Алгоритмы заняли первое место на международном соревновании «dunnhumby's Shopper Challenge»

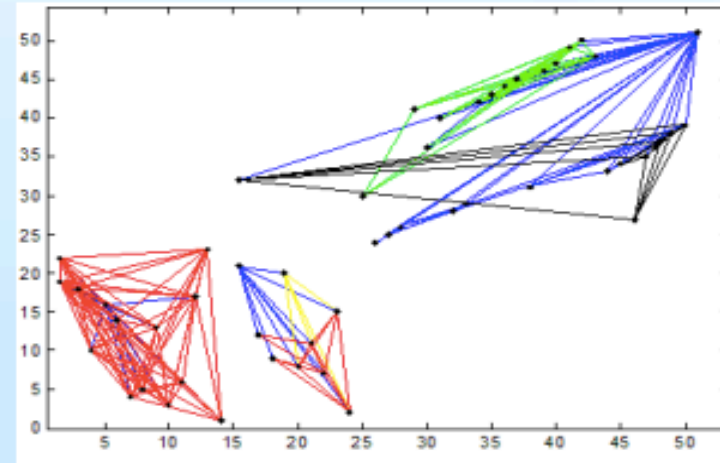


Примеры задач машинного обучения

Алгоритм выявления сообществ в социальных сетях



Матрица смежности фрагмента реальной социальной сети



Граф друзей конкретного пользователя, разбитый на сообщества (показаны цветом)

Дано: граф соцсети (вершины – пользователи, рёбра – отношения дружбы), описания пользователей (ФИО, пол, возраст, и т.п.)

Найти: множества вершин, которые соответствуют реальным группам в соцсети

Метод: матричные разложения матрицы смежности, кластеризация в пространстве главных компонент

Примеры задач машинного обучения

- **Кредитный скоринг:**
объект – данные о заёмщике
ответ – вероятность дефолта, решение по кредиту
- **Предсказание оттока клиентов:**
объект – данные о клиенте на момент времени t
ответ – уйдёт ли клиент к моменту времени $t + \Delta$
- **Прогнозирование объёмов продаж:**
объект – данные о продажах на момент времени t
ответ – объём спроса в интервале от t до $t + \Delta$



Бум искусственного интеллекта

- 1997:** IBM Deep Blue обыграл чемпиона мира по шахматам
- 2005:** Беспилотный автомобиль: DARPA Grand Challenge
- 2006:** Google Translate – статистический машинный перевод
- 2011:** 40 лет DARPA CALO привели к созданию Apple Siri
- 2011:** IBM Watson победил в ТВ-игре «Jeopardy!»
- 2011–2015:** ImageNet: 25% → 3,5% ошибок против 5% у людей
- 2015:** Фонд OpenAI в \$1 млрд. Илона Маска и Сэма Альтмана
- 2016:** DeepMind, OpenAI: динамическое обучение играм Atari
- 2016:** Google DeepMind обыграл чемпиона мира по игре го
- 2017:** OpenAI обыграл чемпиона мира по компьютерной игре Dota 2



Пререквизиты

- Математический анализ 1-2
- Линейная алгебра 1-2
- Математическая статистика

Пример задач, которые предполагается, что слушатели умеют решать:

<http://www.cs.cmu.edu/~ninamf/courses/601sp15/hw/homework1.pdf>

Авторы курса

Андрейцев Антон Игоревич

Bart-69@mail.ru

Бирюкова Валентина Викторовна

mytykritik777@gmail.com