

Прикладные задачи анализа данных

ВВОДНАЯ ЛЕКЦИЯ

Дьяконов А.Г.

**Московский государственный университет
имени М.В. Ломоносова (Москва, Россия)**



История курса

2013 «ПЗАД» спецкурс (лекции + задания)

2014 «ПЗАД» спецкурс (лекции + задания)

2015 «ПЗАД» спецкурс (лекции + задания)

2015 «АМА» Магистратура 1г. ММП ВМК (лекции + семинары)

2016 «ПЗАД» Магистратура 1г. ММП ВМК (лекции + семинары)

2017 «ПЗАД» Магистратура 1г. ММП ВМК (лекции + семинары)

2018 «ПЗАД» Магистратура 1г. ММП ВМК (лекции + семинары)

+ «Введение в машинное обучение» (спецкурс)

+ «Машинное обучение и анализ данных» (поточный курс)

+ «Глубокое обучение» (часть курса)

Цель

Порешать реальные задачи

Обратить внимание на некоторую теорию

Источники

- **Соревновательные платформы**
- **Бизнес**

Задачи

«small data» ~ до 200Гб

Максимально удобно для новичков

Простые форматы данных

Можно решать на компьютере

Лекции

- Разбор кейсов (курс ПЗАД)
- Теория по отдельным темам

Семинары

- Решение задач по теории
- Отчёты по выполнению ДЗ (**публичное выступление**)
 - Обсуждение решений (мозговой штурм)
 - Опросы по темам
- Обучению программированию (**R / Matlab / Python**)

Зачем

- Привлечь к решению задач
- Необходим практикум (текст, звуки, html и т.п.)
 - Обмен опытом
- Разделение труда / стратегии в соревнованиях
 - Написание статей
- Помощь на начальном этапе (код)

Почему стоит заниматься «практикой»

- Любая теория – для решения задач
- Задачника по АД нет!
- Есть возможность решать современные задачи
 - BCI
 - Flickr
 - Предсказание результатов тестирования
- Любая теория – для решения задач
- Объективная оценка исследований

Что даст

- Нет плохих / хороших алгоритмов
- Есть плохие / хорошие решения
- Реальные задачи – не те которым учат

| | |
|--------------------------------|--|
| Фундаментальный уровень | часто нет чистой регрессии, классификации и т.п. Пример: прогноз покупок товара, где 0 – не ноль покупок, а нет на складе |
| Постановочный уровень | просто другие формулировки, другие требования к решению Пример: улучшение на доли процентов |
| Практический уровень | другие объёмы Пример: 50 объектов задачи 70х годов |

Что требуется от слушателей

- **Компьютер (ОЗУ!)**
- **Python, ...**
- **Знания (машинное обучение, статистика и т.д.)**
- **Время!**
- **Внимание (волшебный признак в задаче о страховке)**

Что такое соревнование

Компания – Платформа – Решатели



1. Данные
2. Функционал качества
3. Регламент
4. Система обмена опытом

Что такое соревнование

Правила

- 1. Один ник**
- 2. Нельзя обмениваться кодом/идеями**

...

Наше участие

ник: Ivan Grozniy (MMP, MSU, Russia)

Что такое бизнес-задача

Данные

Неполные

Противоречивые

В разных форматах

Много/мало

Качество

Слабая формализация

Несколько функционалов

Бизнес-термины

Пример: проценты

Сроки

Система внедрения (это решение, которое будет работать!)

Система оценки

серия заданий
по каждому от 0 до 10 штрафных баллов
0 – верно и **в срок**

задание (практическое)

- **качественное** решение (например, в LB + преодоление бенчмарка)
 - код решения (например, выкладывается в форуме)
 - отчёт по решению / слайды (+ делается доклад на семинаре)
- последние два пункта – **для группы**

занятие теоретическое

- контрольные работы
(репетиция экзамена)

бонусы

- до 10 антиштрафных баллов за лучшее решение
- опционально за активность (обсуждение в форуме)

итоговая оценка:

≤ 10 ш.б. – отлично **автоматом**

≤ 20 ш.б. – хорошо

≤ 30 ш.б. – удовлетворительно

> 30 ш.б. – неудовлетворительно **автоматом**

**Экзамен письменный – с коррекцией оценки
(как – позже уточним)**

страница курса

<http://goo.gl/PwWBbr>

<https://github.com/Dyakonov/PZAD>