

ВВЕДЕНИЕ В ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И БИЗНЕС ИНТЕЛЛЕКТ

Аннотация. Курс содержит введение в науку о больших данных и основные методы искусственного интеллекта для их обработки. Наука о данных, искусственный интеллект и бизнес интеллект являются ключевыми элементами цифровой экономики. Поэтому курс рассчитан, на повышение цифровой грамотности управленцев, представителей коммерческих компаний и всех заинтересованных категорий населения. Каждая тема помимо теоретического материала, содержит примеры приложений методов искусственного интеллекта в различных областях экономики, бизнеса, управления и т.д. Курс также будет весьма полезен для повышения квалификации преподавателей и сотрудников университетов и других образовательных учреждений. Общая продолжительность курса составляет примерно 40 ч.

Ключевые слова: Наука о данных, искусственный интеллект, машинное обучение, бизнес интеллект, цифровая экономика

Введение. Почему Наука о Данных и Искусственный Интеллект?

Тема 1. Основные элементы анализа данных (8 ч.)

- 1.1. Подготовка данных
 - Формат данных
 - Типы переменных
 - Выбор переменных
 - Конструирование признаков
 - Неполные данные
- 1.2. Выбор алгоритма
 - Обучение без учителя
 - Обучение с учителем
 - Обучение с подкреплением
 - Другие факторы
- 1.3. Настройка параметров
- 1.4. Оценка результатов
 - Метрики классификации
 - Метрика регрессии
 - Валидация
- 1.5. Обзор темы

Тема 2. Кластеризация методом k-средних и анализ клиентов (2 ч.)

- 2.1. Поиск кластеров клиентов
- 2.2. Пример: профили покупателей
- 2.3. Определение кластеров
 - Сколько кластеров существует?
 - Что включают кластеры?
- 2.4. Ограничения
- 2.5. Обзор темы

Тема 3. Метод главных компонент и анализ товаров (2 ч.)

- 3.1. Изучение товарной ценности
- 3.2. Главные компоненты
- 3.3. Пример: анализ групп товаров
- 3.4. Ограничения
- 3.5. Обзор темы

Тема 4. Ассоциативные правила и анализ продаж (2 ч.)

- 4.1. Поиск покупательских шаблонов
- 4.2. Поддержка, достоверность и лифт
- 4.3. Пример: ведение продуктовых продаж
- 4.4. Принцип A priori
 - Поиск товарных наборов с высокой поддержкой
 - Поиск товарных правил с высокой достоверностью или лифтом
- 4.5. Ограничения
- 4.6. Обзор темы

Тема 5. Анализ социальных сетей и профили потребителей (2 ч.)

- 5.1. Составление схемы отношений
- 5.2. Пример: геополитика в торговле
- 5.3. Лувенский метод
- 5.4. Алгоритм PageRank
- 5.5. Ограничения
- 5.6. Обзор темы

Тема 6. Регрессионный анализ и цены на жилье (2 ч.)

- 6.1. Выведение линии тренда
- 6.2. Пример: предсказание цен на жилье
- 6.3. Градиентный спуск
- 6.4. Коэффициенты регрессии
- 6.5. Коэффициенты корреляции
- 6.6. Ограничения
- 6.7. Обзор темы

Тема 7. Метод k-ближайших соседей и обнаружение аномалий в товарной экспертизе (4 ч.)

- 7.1. Товарная экспертиза
- 7.2. Яблоко от яблони недалеко падает
- 7.3. Пример: истинные различия в вине
- 7.4. Обнаружение аномалий
- 7.5. Ограничения
- 7.6. Обзор темы

Тема 8. Метод опорных векторов и медицинская диагностика (2 ч.)

- 8.1 «Нет» или «о, нет!»?
- 8.2. Пример: обнаружение хронических заболеваний
- 8.3. Построение оптимальной границы
- 8.4. Ограничения
- 8.5. Обзор темы

Тема 9. Дерево решений и прогнозы чрезвычайных ситуаций (2 ч.)

- 9.1. Прогноз выживания в катастрофе
- 9.2. Пример: спасение с тонущего «Титаника»
- 9.3. Создание дерева решений
- 9.4. Ограничения
- 9.5. Обзор темы

Тема 10. Случайные леса и предсказание криминальной активности (2 ч.)

- 10.1. Мудрость толпы
- 10.2. Пример: предсказание криминальной активности
- 10.3. Ансамбли
- 10.4. Бэггинг
- 10.5. Ограничения
- 10.6. Обзор темы

Тема 11. Нейронные сети и создание новых продуктов с заданными свойствами (6 ч.)

- 11.1. Создание мозга
- 11.2. Пример: распознавание рукописных цифр
- 11.3. Компоненты нейронной сети
- 11.4. Правила активации
- 11.5. Ограничения
- 11.6. Обзор темы

Тема 12. A/B-тестирование и многорукие бандиты (2 ч.)

- 12.1. Основы A/B-тестирования
- 12.2. Ограничения A/B-тестирования
- 12.3. Стратегия снижения эpsilon
- 12.4. Пример: многорукие бандиты
- 12.5. Забавный факт: ставка на победителя
- 12.6. Ограничения стратегии снижения эpsilon
- 12.7. Обзор темы

Приложения (4 ч.)

- A. Обзор алгоритмов обучения без учителя
- B. Обзор алгоритмов обучения с учителем
- C. Список параметров настройки
- D. Другие метрики оценки
 - Метрики классификации
 - Метрики регрессии

Итого: 40 ч.

Практические занятия: По каждой теме предполагаются давать в виде домашних работ и самостоятельных проектов

Рекомендуемая литература

Основная:

Бринк Хенрик, Ричардс Джозеф, Феверолф Марк. Машинное обучение. — СПб.: Питер, 2018. — 336 с.: ил.

Винстон Уэйн. Бизнес-моделирование и анализ данных. Решение актуальных задач с помощью Microsoft Excel. 5-е издание. — СПб.: Питер, 2018. — 864 с.: ил.

Николенко С., Кадуринов А., Архангельская Е. Глубокое обучение. — СПб.: Питер, 2018. — 480 с.: ил.

Плас Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с.: ил.

Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и наука о данных. — СПб.: Питер, 2018. — 336 с.: ил.

Феррейра Фило Владстон. Теоретический минимум по Computer Science. Все, что нужно программисту и разработчику. — СПб.: Питер, 2019. — 224 с.: ил.

Шолле Франсуа. Глубокое обучение на Python и R. — СПб.: Питер, 2018. — 400 с.: ил.

Дополнительная:

Kosinski, M., Matz, S., Gosling, S., Popov, V., & Stillwell, D. (2015). Facebook as a Social Science Research Tool: Opportunities, Challenges, Ethical Considerations and Practical Guidelines. *American Psychologist*.

Agricultural Research Service, United States Department of Agriculture (2015). USDA Food Composition Databases.

Dataset is included in the following R package: Hahsler, M., Buchta, C., Gruen, B., & Hornik, K. (2016).

Stockholm International Peace Research Institute (2015).

Harrison, D., & Rubinfeld, D. (1993). Boston Housing Data.

Forina, M., et al. (1998). Wine Recognition Data.

Robert Detrano (M.D., Ph.D), from Virginia Medical Center, Long Beach and Cleveland Clinic Foundation (1988). Heart Disease Database (Cleveland).

British Board of Trade Inquiry (1990). Titanic Data.

SF OpenData, City and County of San Francisco (2016). Crime Incidents.

National Oceanic and Atmospheric Administration, National Centers for Environmental Information (2016). Quality Controlled Local Climatological Data (QCLCD).

LeCun, Y., & Cortes, C. (1998). The MNIST Database of Handwritten Digits.

Преподаватель: проф. Сергей Павлович Левашкин, заведующий Научно-исследовательской Лабораторией Искусственного Интеллекта ПГУТИ, действительный член Мексиканской Академии Наук, выпускник МГУ им. М.В. Ломоносова, ученый с более чем 20-летним опытом работы в университетах и компаниях России, Северной Америки и Европы levashkin.com