

Программа курса

Лекция 1. Входные требования

- Постановки задач
- Обзор программы
- Термины линейной алгебры. Скалярное произведение.
- Примеры представления объектов
- Правила работы с матрицами и векторами на уровне 1-2 курса..
- Собственные значения и вектора матрицы

Лекция 2. Основные типы моделей для поиска закономерностей в данных

- Регрессионный анализ
- Классификация данных
- Кластеризация данных
- Простые и обобщенные деревья решений
- Ансамбли алгоритмов. Бустинг.

Лекция 3. Модели для поиска закономерностей в данных

- Сокращение данных – метод главных компонент
- Эволюционные алгоритмы
- Нейронные сети. Архитектура нейросетей.
- Задачи, решаемые нейросетями.

Лекция 4. Введение в анализ данных

- Анализ данных и распознавание образов
- Статистические распределения.
- Нормальное распределение
- Гистограмма признака и ее построение.
- Первичное преобразование данных, поиск выбросов

Лекция 5. Введение в анализ данных

- Регрессионный анализ, скользящий контроль
- Деревья решений, простая и обобщенная формы
- Логистическая регрессия

Лекция 6 Близость (похожесть) объектов. Кластеры и их поиск

- Кластер как связная компонента графа.
- Волновой алгоритм. Алгоритм DBSCAN.
- Построение минимального покрывающего дерева.

Лекция 7 Кластеры и их поиск

- Метод К средних, Классическая версия. Версия с ядрами.
- Нечеткая принадлежность к кластеру. EM-алгоритм.
- Иерархический кластер-анализ, дендрограмма.
- Формальный элемент. Метод ФорЭль.

Лекция 8. Метод главных компонент

- Скрытые переменные.
- Факторы и их поиск.
- SVD разложение матрицы.

Лекция 9. Факторы. Метод главных компонент

- Геометрический смысл факторов.
- Регрессия на факторах
- Вычислительная химия. Метод COMFA.
- Многомерное шкалирование

Лекция 10. Продвинутое методы анализа

- Эволюционный алгоритм – МГУА. Линейная версия.
- Поиск подмножеств признаков в регрессии.
- Нелинейное преобразование признаков.
- Эволюционный алгоритм – генетический.
- Гены и фенотип.
- Поиск подмножеств признаков для вычисления метрики.

Лекция 11. Продвинутое методы анализа.

- Кernels функции – «беспризнаковый» анализ данных
- SVM и поддерживающие вектора

Лекция 12. Продвинутое методы анализа.

- «Когда мало данных» – Метод Bootstrap
- Семейства прогнозирующих алгоритмов
- «Нечеткие» признаки (Fuzzy)
- «Нечеткие» классификаторы

Лекция 13. Нейронные сети. Часть 1

- Модель персептрона и ее ограничения
- Классические нейронные сети, слой нейронов, два типа нейронов
- Задачи, решаемые нейронными сетями, «Глубокое обучение»

Лекция 14. Нейронные сети. Часть 2

- Анализ Изображений и сверточные нейросети
- Нейросети и «Инженерия признаков»
- Проблема переобучения
- Перспективы развития нейросетей
- Графические процессоры (GPU).

Лекция 15. Закрепление знаний

- Повторение основного материала на проходном примере
- Молекулярные графы и их описания.
- Задача «Структура-свойство»
- Подведение итогов