

План курса

«Технологии среды R и Python»

Трек «Технологии анализа экономических данных»

5 семестр

Эта часть курса посвящена применению технологий программных сред R и Python в прикладном интеллектуальном анализе экономических данных.

Занятие 1. Введение

- 1.1. Data Mining как направление анализа данных
- 1.2. Статистические среды R и Python и их использование в Data Mining

Занятие 2. Статистические модели: критерии и методы оценивания их качества (Ч1)

- 2.1. Основные шаги построения и верификации моделей
- 2.2. Использование алгоритмов ресэмплинга для тестирования моделей и оптимизации их параметров
- 2.3. Модели для предсказания класса объектов

Занятие 3. Статистические модели: критерии и методы оценивания их качества (Ч2)

- 2.4. Проецирование многомерных данных на плоскости
- 2.5. Многомерный статистический анализ данных
- 2.6. Методы кластеризации

Занятие 4. Подготовка данных для статистических моделей

- 3.1. Интерфейс доступа к функциям машинного обучения
- 3.2. Обнаружение и удаление «ненужных» предикторов
- 3.3. Предварительная обработка: преобразование и групповая трансформация переменных
- 3.4. Заполнение пропущенных значений в данных

Занятие 5. Построение регрессионных моделей различного типа (Ч1)

- 4.1. Селекция оптимального набора предикторов линейной модели
- 4.2. Регуляризация, частные наименьшие квадраты и kNN-регрессия
- 4.3. Построение деревьев регрессии

Кейс: прогнозирование котировок активов на бирже

Занятие 6. Построение регрессионных моделей различного типа (Ч2)

- 4.4. Ансамбли моделей: бэггинг, случайные леса, бустинг
- 4.5. Сравнение построенных моделей и оценка информативности предикторов
- 4.6. Деревья регрессии с многомерным откликом

Кейс: прогнозирование котировок активов на бирже

Занятие 7. Бинарные матрицы и ассоциативные правила

- 5.1. Классификация в бинарных пространствах с использованием классических моделей

- 5.2 Бинарные деревья решений
- 5.3 Поиск логических закономерностей в данных
- 5.4 Алгоритмы выделения ассоциативных правил
- 5.5 Анализ последовательностей знаков или событий

Кейс: анализ потребительской корзины

Занятие 8. Бинарные классификаторы с различными разделяющими поверхностями (Ч1)

- 6.1 Дискриминантный анализ
- 6.2 Метод опорных векторов
- 6.3 Ядерные функции машины опорных векторов

Кейс: обнаружение факта мошенничества по базе транзакций по кредитным картам

Занятие 9. Бинарные классификаторы с различными разделяющими поверхностями (Ч2)

- 6.4 Деревья классификации, случайный лес и логистическая регрессия
- 6.5 Процедуры сравнения эффективности моделей классификации

Кейс: обнаружение факта мошенничества по базе транзакций по кредитным картам

Занятие 10. Модели классификации для нескольких классов (Ч1)

- 7.1 Ирисы Фишера и метод k ближайших соседей
- 7.2 Наивный байесовский классификатор
- 7.3 Классификация в линейном дискриминантном пространстве

Кейс: решение задач кредитного скоринга

Занятие 11. Модели классификации для нескольких классов (Ч2)

- 7.4 Нелинейные классификаторы
- 7.5 Модель мультиномиального логита
- 7.6 Классификаторы на основе искусственных нейронных сетей

Кейс: решение задач кредитного скоринга

Занятие 12. Моделирование порядковых и счетных переменных (Ч1)

- 8.1 Модель логита для порядковой переменной
- 8.2 Настройка параметров нейронных сетей
- 8.3 Методы комплексации модельных прогнозов

Занятие 13. Моделирование порядковых и счетных переменных (Ч2)

- 8.4 Обобщенные линейные модели для счетных данных
- 8.5 ZIP- и барьерные модели счетных данных

Занятие 14. Методы многомерной ординации

- 9.1 Преобразование данных и вычисление матрицы расстояний
- 9.2 Непараметрический дисперсионный анализ матриц дистанций
- 9.3 Методы ординации объектов и переменных: построение и сравнение диаграмм
- 9.4 Оценка связи ординации с внешними факторами
- 9.5 Неметрическое многомерное шкалирование и построение распределения чувствительности видов

Занятие 15. Кластерный анализ (Ч1)

10.1 Алгоритмы кластеризации, основанные на разделении

10.2 Иерархическая кластеризация

10.3 Оценка качества кластеризации

Кейс: кластеризация регионов РФ по социально-экономическим показателям

Занятие 16. Кластерный анализ (Ч2)

10.4 Другие алгоритмы кластеризации

10.5 Самоорганизующиеся карты Кохонена

Кейс: кластеризация регионов РФ по социально-экономическим показателям