Язык R в примерах и задачах

Юрченков Иван Александрович

14 09 2020

Table of Contents

# 1 Основы программирования на языке R

## 1.1 Введение

## 1.2 Переменные. Типы данных. Приведение типов в R

## 1.3 Векторы как элементарная структура данных

## 1.4 Управляющие конструкции и циклы

## 1.5 Функции

## 1.6 Матрицы

## 1.7 Массивы

## 1.8 Строки. Работа со строковыми переменными

## 1.9 Списки

## 1.10 Обработка исключений

## 1.11 Работа с датами в R

## 1.12 Подключаемые библиотеки

# 2 Работа с данными в R

## 2.\* Виды данных

## 2.\* Работа с фреймами данных. Агрегация

## 2.\* Статистические критерии. Статистический тест и стандартные распределения

## 2.\* Библиотеки для работы с данными

## 2.\* Дополнительные библиотеки фреймов

## 2.\* Подключение к sql базе данных. Считывание и запись данных

# 3 Визуализация данных в R

## 3.1 Стандартная визуализация с использованием plot

## 3.2 Стандартная визуализация объемных данных

## 3.3 Визуализация данных с использованием ggplot2

## 3.4 Визуализация данных с использованием plotly

## 3.5 Визуализация данных на картах

# 4 Машинное обучение с R

## 4.1 Задачи машинного обучения

## 4.2 Линейная регрессия

## 4.\* О выборках, мерах качества и кросс-валидации

## 4.3 Многокритериальная регрессия

## 4.4 Нелинейная регрессия

## 4.5 Логистическая регрессия. Множественная классификация

## 4.6 SVM. Метод опорных векторов

## 4.7 Наивный Байесовский классификатор

## 4.8 Линейный дискриминантный анализ

## 4.9 Метод k-ближайших соседей

## 4.10 Метод k-средних

## 4.11 Другие методы кластеризации

## 4.12 Дерево решений для классификации и регрессии

## 4.13 Случайный лес

## 4.14 Алгоритм градиентного бустинга

# 5 Методы глубокого обучения

## 5.1 Нейроны и нейронные сети

## 5.2 Полносвязная нейронная сеть

## 5.3 Свёрточные нейронные сети

## 5.4 Рекуррентные нейронные сети

## 5.5 Механизм внимания и трансформеры

# 6 Анализ данных временных рядов

## 6.\* Основные компоненты временного ряда данных

## 6.\* Статистические закономерности временного ряда. Стационарность, сезонность

## 6.\* Выделение компонент временного ряда

## 6.\* Механизм авторегрессии

## 6.\* Ансамблевые модели временных рядов

# 7 Предиктивный анализ данных

## 7.\* Постановка проблемы реального распределения в данных

## 7.\* Эвристические алгоритмы поиска реального распределения. Анаморфозы реальных распределений

## 7.\* Анализ медленных движений временных рядов. Модели роста

## 7.\* Анализ быстрых движений временного ряда. Колебания данных относительно тренда

## 7.\* Связь медленных и быстрых компонент ряда данных

## 7.\* Статистические методы анализа многомерных структур. Структура объектов со множественными характеристиками

## 7.\*

# Доп Математика

## Доп.\* Основы линейной алгебры

## Доп.\* Матричные алгоритмы

## Доп.\* Математический анализ

## Доп.\* Теория вероятностей и математическая статистика

## Доп.\* Теория информации

# Темы курсовых работ

## Программирование на языке R

1. Реализация алгоритма многопараметрической линейной регрессии. Решение задач прогнозирования
2. Реализация алгоритма многоклассового классификационного метода на основе логистической регрессии
3. Реализация алгоритма Байесовской классификации
4. Реализация алгоритма кластеризации методом к-средних
5. Реализация EM-алгоритма кластеризации
6. Реализация алгоритмов уменьшения размерности данных. Метод главных компонент
7. Реализация алгоритмов уменьшения размерности данных. Метод независимых компонент
8. Реализация алгоритмов уменьшения размерности данных. Нейронная сеть автокодировщик
9. Факторный анализ. Реализация методов Фишера для линейного дискриминантного анализа
10. Реализация нейронной сети прямого распространения в задачах регрессионного анализа
11. Реализация нейронной сети прямого распространения в задачах прогнозирования временных рядов
12. Реализация методов авторегрессии
13. Реализация алгоритма дерева решений. Дерево решений в задачах классификации
14. Реализация алгоритма дерева решений. Дерево решений в задачах регрессии
15. Реализация алгоритма случайного леса. Случайный лес в задачах регрессионного анализа
16. Реализация алгоритмов решения операторных уравнений для матриц высокой размерности.

## Анализ данных в R

1. Прогнозирование временных рядов. Прогнозирование средних доходов населения России по отношению к доходам федерального бюджета. Заработная плата: <http://sophist.hse.ru/hse/1/tables/WAG_M.htm>, доходы федерального бюджета: <http://sophist.hse.ru/hse/1/tables/GOV_M.htm>
2. Прогнозирование временных рядов. Прогнозирование цен акций мировых компаний: <https://finance.yahoo.com>
3. Исследование влияние переобучения в моделях регрессии и прогнозирование на основе регрессионных моделей с использованием регуляризации.