

Миркин, Б. Г.

Введение в анализ данных : учебник и практикум для бакалавриата и магистратуры /
Б. Г. Миркин. — М.: Издательство Юрайт, 2016. — 174 с. — Серия : Авторский учебник.

Оглавление

Предисловие.....	5
Глава 1. Что такое анализ данных?	11
1.1. Иллюстративные проблемы анализа данных.....	11
1.2. Анализ данных и математическая статистика: инженерный и научный подходы...	19
Глава 2. Одномерный анализ.....	27
2.1. Количественные признаки: распределение и гистограмма.....	28
П2.1. Представление.....	28
Ф2.1. Формулировки.....	30
В2.1. Вычисление.....	32
2.2. Дальнейшая суммаризация: центр и рассеяние.....	33
П2.2. Центр и рассеяние: представление.....	33
Ф2.2. Центр и рассеяние: формулировки.....	36
Ф2.2.1. Подход анализа данных.....	36
Ф2.2.2. Теоретико-вероятностный подход.....	39
В2.2. Центр и рассеяние: вычисление.....	41
2.3. Бинарные и категоризованные признаки.....	41
П2.3. Представление.....	41
Ф2.3. Формулировки.....	45
В2.3. Вычисление.....	48
Проект 2.1. Вычисление центра по критерию Минковского.....	48
Проект 2.2. Доверительный интервал бутстрэп-методом.....	50
Проект 2.3. Перекрестная валидация (скользящий контроль).....	54
Глава 3. Двумерный анализ: суммаризация и корреляция двух признаков. . .	59
3.1. Постановка проблемы.....	60
3.2. Случай двух количественных признаков.....	60
ПЗ.2. Линейная регрессия: представление.....	60
ПЗ.2.1. Поле рассеяния, линейная регрессия и коэффициент корреляции... ..	60
ПЗ.2.2. Анализ степени адекватности уравнения регрессии.....	63
ФЗ.2. Линейная регрессия: формулировки.....	69
ФЗ.2.1. Аппроксимация данных линейным уравнением.....	69
ФЗ.2.2. Коэффициент корреляции и его свойства.....	70
ФЗ.2.3. Линеаризация для оценки нелинейной регрессии.....	72
ВЗ.2. Линейная регрессия: вычисление.....	72
Проект 3.1. Линейная регрессия и бутстрэп.....	73
Проект 3.2. Нелинейная и линеаризованная регрессии: инспирированный природой алгоритм.....	77
3.3. Случай смешанных шкал: номинальный и количественный признаки.....	83
3.3.1. Целевой количественный признак.....	83
ПЗ.3.1. Бокс-плот, табличная регрессия и корреляционное отношение.....	83
ФЗ.3.1. Табличная регрессия: формулировки.....	87

3.3.2. Номинальный целевой признак.....	89
3.3.2.1. Классификатор по правилу ближнего соседа.....	89
3.3.2.2. Классификатор с интервальными предикатами.....	92
3.4. Случай двух номинальных признаков.....	94
ПЗ.4, Анализ таблиц сопряженности: представление.....	94
ПЗ.4.1, Построение концептуальных связей по статистическим данным ...	94
ПЗ.4.2. Исследование связей с помощью индекса Кетле.....	97
ПЗ.4.3. Коэффициент хи-квадрат как индекс связи и визуализация его структуры.....	102
ФЗ.4. Анализ таблиц сопряженности: формулировки.....	105
Глава 4. Корреляция и суммаризация для многомерных данных.....	110
4.1. Необходимость задания класса решающих правил при изучении корреляции ...	111
4.2. Байесовский подход к распознаванию.....	115
4.2.1. Байесовское решающее правило.....	115
4.2.2. Наивный Байесовский классификатор.....	117
4.3. Меры качества классификатора ...	120
П4.3. Точность и связанные с ней показатели.....	120
Ф4.3. Точность и связанные с ней показатели: формулировки.....	122
4.4. Постановка проблемы кластеризации.....	124
4.5. Кластеризация методом iC -средних.....	127
П4.5< Параллельный метод iC -средних и его особенности.....	127
Ф4.5, Критерий метода X -средних.....	134
В4.5. Вычисления по методу It -средних с использованием системы МатЛаб.....	136
4.6. Проблема инициализации \wedge -средних и аномальные кластеры.....	138
4.6.1. Подходы к инициализации iC -средних.....	138
4.6.1.1. Многочисленные прогоны для инициализации IC -средних.....	138
4.6.1.2. Метод аномальных кластеров.....	140
П4.6.1.2. Аномальные кластеры.....	140
Ф4.6.1.2. Аномальные группы.....	143
4.6.2. Интеллектуальная версия метода X -средних.....	145
П4.6.2. Аномальные группы и интеллектуальный метод iC -средних.....	145
Ф4.6.2. Метод iK -средних: формулировки и вычисление.....	146
Проект 4.1. Роль предварительного преобразования по методу главных компонент ...	149
Заключение.....	154
Приложение.....	158
Предметный указатель.....	171
Список литературы.....	173