

Анналин Ын, Кеннет Су

Научная библиотека

БНТУ



★ 8 0 1 2 3 4 1 5 1 ★

ТЕОРЕТИЧЕСКИЙ МИНИМУМ ПО BIG DATA

ВСЁ, ЧТО НУЖНО ЗНАТЬ О БОЛЬШИХ ДАННЫХ

НАУКОВАЯ БІБЛІЯТЭКА

Беларускага нацыянальнага
тэхнічнага ўніверсітэта

Інв. № 1883848



Санкт-Петербург • Москва • Минск

2022

Краткое содержание

Метрики классификации.....	34
Метрика регрессии.....	35
Выводы.....	35
Предисловие.....	12
Введение.....	16
Почему Data Science?.....	18
Глава 1. Об основах без лишних слов.....	21
Глава 2. Кластеризация методом k-средних.....	39
Глава 3. Метод главных компонент.....	51
Глава 4. Ассоциативные правила.....	65
Глава 5. Анализ социальных сетей.....	77
Глава 6. Регрессионный анализ.....	93
Глава 7. Метод k-ближайших соседей и обнаружение аномалий.....	107
Глава 8. Метод опорных векторов.....	117
Глава 9. Дерево решений.....	127
Глава 10. Случайные леса.....	137
Глава 11. Нейронные сети.....	149
Глава 12. A/B-тестирование и многорукие бандиты.....	167
Приложения.....	179
Глоссарий.....	188
Литература и ссылки на источники.....	199
Об авторах.....	204

Оглавление

Предисловие	12
От издательства.....	15
Введение	16
Почему Data Science?	18
Глава 1. Об основах без лишних слов.....	21
1.1. Подготовка данных.....	22
Формат данных.....	23
Типы переменных	24
Выбор переменных	25
Конструирование признаков	25
Неполные данные	26
1.2. Выбор алгоритма.....	27
Обучение без учителя.....	28
Обучение с учителем	29
Обучение с подкреплением	30
Другие факторы.....	31

1.3. Настройка параметров	31
1.4. Оценка результатов	33
Метрики классификации	34
Метрика регрессии	35
Валидация	36
1.5. Краткие итоги	38
Глава 2. Кластеризация методом k-средних	39
2.1. Поиск кластеров клиентов	40
2.2. Пример: профили кинозрителей	41
2.3. Определение кластеров	42
Сколько кластеров существует?	44
Что включают кластеры?	46
2.4. Ограничения	48
2.5. Краткие итоги	49
Глава 3. Метод главных компонент	51
3.1. Изучение пищевой ценности	52
3.2. Главные компоненты	53
3.3. Пример: анализ пищевых групп	56
3.4. Ограничения	61
3.5. Краткие итоги	64
Глава 4. Ассоциативные правила	65
4.1. Поиск покупательских шаблонов	66
4.2. Поддержка, достоверность и лифт	67

4.3. Пример: ведение продуктовых продаж	69
4.4. Принцип Argiоri	72
Поиск товарных наборов с высокой поддержкой.....	73
Поиск товарных правил с высокой достоверностью или лифтом.....	74
4.5. Ограничения	75
4.6. Краткие итоги.....	76
Глава 5. Анализ социальных сетей	77
5.1. Составление схемы отношений.....	78
5.2. Пример: геополитика в торговле оружием.....	80
5.3. Лувенский метод	84
5.4. Алгоритм PageRank	86
5.5. Ограничения	90
5.6. Краткие итоги	91
Глава 6. Регрессионный анализ	93
6.1. Выведение линии тренда.....	94
6.2. Пример: предсказание цен на дома	95
6.3. Градиентный спуск	98
6.4. Коэффициенты регрессии.....	101
6.5. Коэффициенты корреляции.....	102
6.6. Ограничения	104
6.7. Краткие итоги.....	106
Глава 7. Метод k-ближайших соседей и обнаружение аномалий	107
7.1. Пищевая экспертиза.....	108

7.2. Яблоко от яблони недалеко падает	109
7.3. Пример: истинные различия в вине	111
7.4. Обнаружение аномалий.....	113
7.5. Ограничения	114
7.6. Краткие итоги.....	115
Глава 8. Метод опорных векторов.....	117
8.1 «Нет» или «о, нет!»?.....	118
8.2. Пример: обнаружение сердечно-сосудистых заболеваний	118
8.3. Построение оптимальной границы.....	120
8.4. Ограничения	124
8.5. Краткие итоги.....	125
Глава 9. Дерево решений.....	127
9.1. Прогноз выживания в катастрофе	128
9.2. Пример: спасение с тонущего «Титаника»	128
9.3. Создание дерева решений	131
9.4. Ограничения	133
9.5. Краткие итоги.....	135
Глава 10. Случайные леса.....	137
10.1. Мудрость толпы	138
10.2. Пример: предсказание криминальной активности.....	139
10.3. Ансамбли	144
10.4. Бэггинг.....	145

10.5. Ограничения.....	147
10.6. Краткие итоги	148
Глава 11. Нейронные сети	149
11.1. Создание мозга	150
11.2. Пример: распознавание рукописных цифр.....	152
11.3. Компоненты нейронной сети.....	156
11.4. Правила активации	159
11.5. Ограничения.....	161
11.6. Краткие итоги	165
Глава 12. А/В-тестирование и многорукие бандиты	167
12.1. Основы А/В-тестирования.....	168
12.2. Ограничения А/В-тестирования.....	169
12.3. Стратегия снижения эpsilon.....	169
12.4. Пример: многорукие бандиты	171
12.5. Забавный факт: ставка на победителя.....	174
12.6. Ограничения стратегии снижения эpsilon.....	175
12.7. Краткие итоги	176
Приложения	179
Приложение А. Обзор алгоритмов обучения без учителя	180
Приложение В. Обзор алгоритмов обучения с учителем	181
Приложение С. Список параметров настройки.....	182

Приложение D. Другие метрики оценки.....	183
Метрики классификации.....	183
Метрики регрессии.....	186
Глоссарий	188
Литература и ссылки на источники	199
Источники на английском языке.....	199
Литература на русском языке.....	202
Об авторах.....	204