

УДК 519.2(075.8)
Т 415

Рецензенты:

д-р техн. наук, профессор *А.А. Попов*,
канд. техн. наук, доцент *В.Ю. Щеколдин*

Работа подготовлена на кафедре теоретической и прикладной информатики для студентов, обучающихся по направлениям подготовки 01.03.02 «Прикладная математика и информатика», 02.03.03 «Математическое обеспечение и администрирование информационных систем»

Тимофеева А.Ю.

Т 415 Вероятностные основы методов и алгоритмов анализа данных : учебное пособие / А.Ю. Тимофеева, Е.А. Хайленко. – Новосибирск : Изд-во НГТУ, 2020. – 82 с.

ISBN 978-5-7782-4166-4

Учебное пособие представляет собой дополнительные материалы по курсу «Статистические методы анализа данных». В пособии приведены вероятностные постановки, описаны модели бинарного отклика и модели с ошибками в переменных, рассмотрены методы и алгоритмы построения регрессионных зависимостей для этих моделей. Кроме того, в пособие включены скрипты на языке R для реализации некоторых расчетов и вывода графиков.

УДК 519.2(075.8)

ISBN 978-5-7782-4166-4

© Тимофеева А.Ю., Хайленко Е.А., 2020
© Новосибирский государственный
технический университет, 2020

ОГЛАВЛЕНИЕ

Введение	4
1. Вероятностные постановки задач анализа данных.....	5
1.1. Условная вероятность и взаимная информация	6
1.2. Функция правдоподобия и метод максимального правдоподобия	14
1.3. Корреляция	21
1.4. Регрессия как условное математическое ожидание	26
1.5. Модели бинарного отклика	29
1.6. Модели с ошибками в переменных	33
1.7. Неоднородные данные	35
2. Алгоритмы анализа данных	39
2.1. ЕМ-алгоритм разделения смесей	39
2.2. Алгоритмы оценивания параметров моделей бинарного отклика	43
2.3. Итерационный алгоритм поиска ММП-оценок параметров моделей с ошибками в переменных	49
3. Методы анализа данных	51
3.1. Ортогональная регрессия	51
3.2. Метод наименьших квадратов синусов	54
3.3. Метод устойчивой средней геометрической регрессии	59
3.4. Метод смешанной устойчивой регрессии	60
3.5. Метод общих наименьших квадратов	69
3.6. Метод скорректированных наименьших квадратов	75
Заключение	78
Библиографический список	79