

1573791

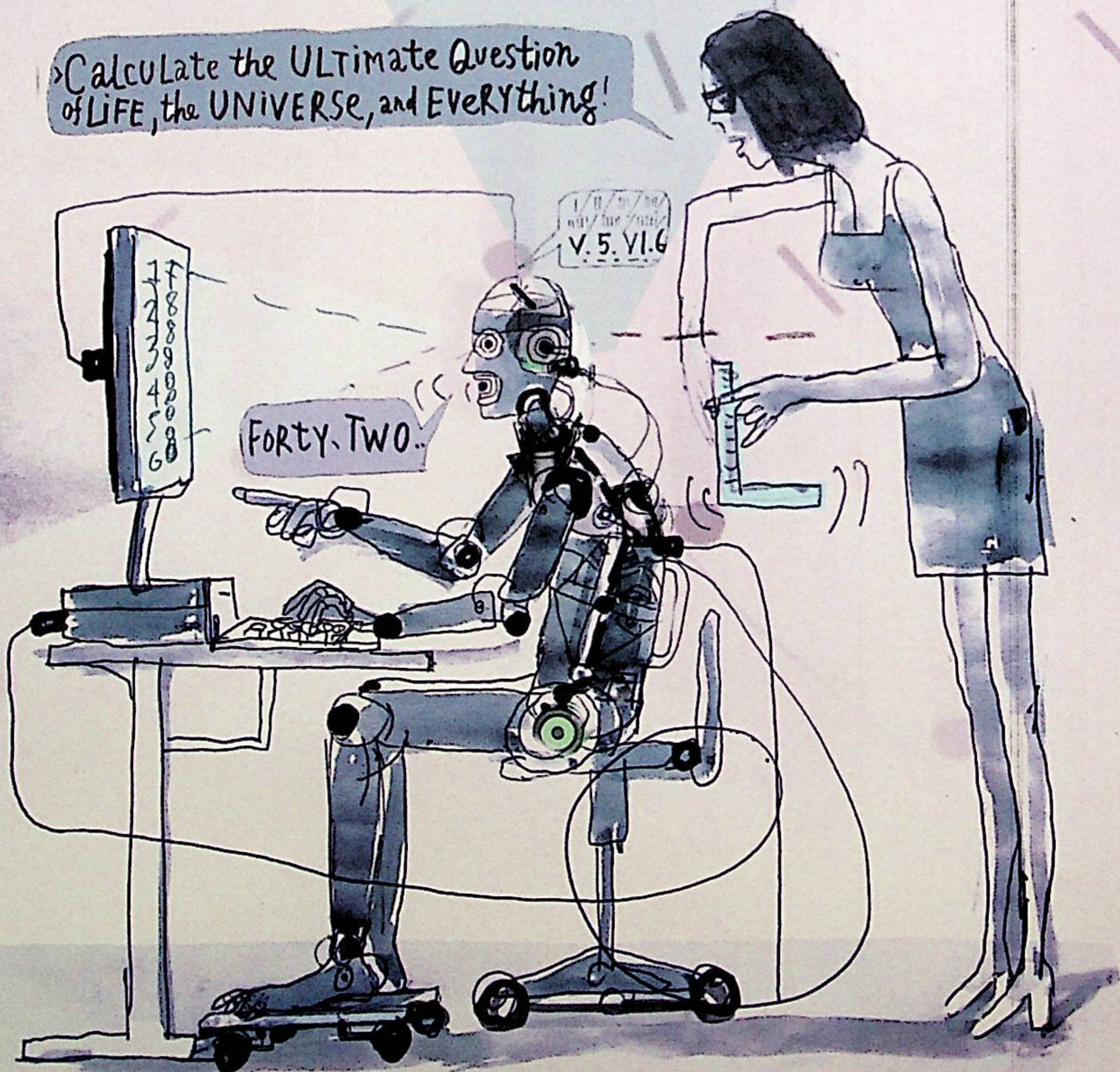
Харківський
національний
університет
імені В. Н. Каразіна



К. Ю. КОНОНОВА

МАШИННЕ НАВЧАННЯ

МЕТОДИ ТА МОДЕЛІ



У підручнику розглядаються основні принципи аналізу даних та машинного навчання, приділяється увага CRISP-DM методології та питанням підготовки даних. Розглянуто базові моделі лінійної та нелінійної регресії, класифікації (зокрема логістична регресія, методи опорних векторів та k-найближчих сусідів, Байєсова класифікація, дерева рішень і випадковий ліс), кластеризації (ієрархічна і k-середніх), а також методи побудови асоціативних правил. Серед більш складних методів машинного навчання розглянуто методи обробки природної мови (модель «Мішок слів», класифікація текстів та аналіз настроїв), використання штучних нейронних мереж в задачах прогнозування, класифікації та кластеризації, подано алгоритми глибокого навчання (згорткові та рекурентні мережі).

Кожен розділ підручника містить теоретичний матеріал, ілюстративні матеріали та лабораторні роботи, виконання яких дозволить студентам оволодіти методами та моделями машинного навчання, а також питання для самоперевірки та завдання для самостійного виконання.

Для бакалаврів, магістрів та докторів філософії, а також викладачів ВНЗ. Підручник буде корисним вченим і фахівцям, які використовують в роботі технології машинного навчання.

ЗМІСТ

ВСТУП	5
РОЗДІЛ 1. Основні поняття інтелектуального аналізу даних	8
Тема 1. Завдання аналізу даних	8
Лабораторна робота 1	12
Тема 2. CRISP-DM методологія	13
Тема 3. Підготовка даних	19
Лабораторна робота 2	23
Питання для самоперевірки	30
Самостійна робота 1	30
РОЗДІЛ 2. Регресія	31
Тема 4. Лінійна регресія	31
Лабораторна робота 3	32
Тема 5. Дерева рішень і випадковий ліс	42
Лабораторна робота 4	44
Питання для самоперевірки	55
Самостійна робота 2	55
РОЗДІЛ 3. Класифікація	56
Тема 6. Логістична регресія	56
Лабораторна робота 5	60
Тема 7. Метод опорних векторів	71
Лабораторна робота 6	75
Тема 8. Метод k-найближчих сусідів	83
Лабораторна робота 7	84
Тема 9. Байєсова класифікація	88
Лабораторна робота 8	91
Тема 10. Дерева рішень та ансамблеві методи	95
Лабораторна робота 9	96
Питання для самоперевірки	108
Самостійна робота 3	108
РОЗДІЛ 4. Кластеризація	109
Тема 11. Ієрархічна кластеризація	109
Лабораторна робота 10	110
Тема 12. Кластеризація на основі k-means	116
Лабораторна робота 11	119
Питання для самоперевірки	124
Самостійна робота 4	124

РОЗДІЛ 5. Асоціативні правила	125
Тема 13. Побудова асоціативних правил	125
Тема 14. Алгоритми APRIORI та ECLAT	128
Лабораторна робота 12	135
Питання для самоперевірки	142
Самостійна робота 5	142
РОЗДІЛ 6. Обробка природної мови	143
Тема 15. Модель «мішок слів»	143
Лабораторна робота 13	145
Тема 16. Моделі з урахуванням семантики	150
Питання для самоперевірки	151
Самостійна робота 6	151
РОЗДІЛ 7. Штучні нейронні мережі	152
Тема 17. Теоретичні основи нейронних мереж	152
Тема 18. НМ в задачах апроксимації та прогнозування	158
А. НМ в задачах апроксимації	158
Лабораторна робота 14	159
Б. НМ в задачах прогнозування	169
Лабораторна робота 15	170
Тема 19. НМ в задачах класифікації	177
Лабораторна робота 16	177
Тема 20. НМ в задачах кластеризації: карти Кохонена	184
Лабораторна робота 17	185
Питання для самоперевірки	193
Самостійна робота 7	194
РОЗДІЛ 8. Глибоке навчання	195
Тема 21. Рекурентні нейронні мережі	196
Лабораторна робота 18	199
Тема 22. Згорткові нейронні мережі	205
Лабораторна робота 19	208
Питання для самоперевірки	210
Самостійна робота 8	211
ДОДАТКИ	212
ЛІТЕРАТУРА	275