

Оглавление

Предисловие	6
Вступление	11
Аудитория	11
Подход к изложению материала	12
Структура книги	13
Благодарности	14
Об авторе	16

Часть I. Первые шаги

Глава 1. Поговорим о процессе обучения	19
1.1. Добро пожаловать	19
1.2. Сфера применения, терминология, прогнозирование и данные	20
1.3. Включение машины в процесс машинного обучения	25
1.4. Примеры обучающихся систем	27
1.5. Оценка обучающихся систем	30
1.6. Процесс создания обучающихся систем	33
1.7. Допущения и реальность обучения	36
1.8. Подведение итогов	38
Глава 2. Основные технические сведения	41
2.1. Настройка среды программирования	41
2.2. Потребность в математическом языке	41
2.3. Программное обеспечение для машинного обучения	43
2.4. Вероятность	43
2.5. Линейные комбинации, взвешенные суммы и скалярные произведения	52
2.6. Геометрический взгляд на вещи: точки в пространстве	61
2.7. Нотация и прием «плюс один»	71
2.8. Добавление крутизны, избавление от смиренной рубашки и нелинейность	73
2.9. NumPy против «всей этой математики»	75
2.10. Проблемы с плавающей запятой	82
2.11. Подведение итогов	84
Глава 3. Предсказание категорий: знакомство с классификацией	86
3.1. Задачи классификации	86
3.2. Простой набор данных для задач классификации	87

3.3. Тренировка и тестирование: избегайте натаскивания	90
3.4. Оценка: подсчет результатов экзамена	94
3.5. Простой классификатор № 1: ближайшие соседи, отношения на расстоянии и допущения	96
3.6. Простой классификатор № 2: наивный байесовский алгоритм, вероятность и нарушенные обещания	102
3.7. Упрощенная оценка классификаторов	105
3.8. Подведение итогов	119
Глава 4. Предсказание числовых значений: знакомство с регрессией	124
4.1. Простой набор данных для задач регрессии	124
4.2. Регрессия методом k -ближайших соседей и сводная статистика	127
4.3. Линейная регрессия и ошибки	132
4.4. Оптимизация: выбор лучшего ответа	139
4.5. Простая оценка и сравнение регрессоров	144
4.6. Подведение итогов	148
 Часть II. Оценка 	
Глава 5. Оценка и сравнение обучающихся систем	153
5.1. Оценка и почему меньше значит больше	153
5.2. Терминология для описания этапов обучения	155
5.3. Майор Том, что-то не так: переобучение и недообучение	163
5.4. От ошибок к затратам	172
5.5. (Повторная) выборка: получение большего за счет меньшего	176
5.6. Разложение ошибки на смещение и дисперсию	195
5.7. Графический метод оценки и сравнения	204
5.8. Сравнение моделей с помощью перекрестной проверки	209
5.9. Подведение итогов	211
Глава 6. Оценка классификаторов	215
6.1. Базовые классификаторы	215
6.2. Помимо точности: показатели эффективности классификации	218
6.3. ROC-кривые	229
6.4. Еще один метод многоклассовой классификации: один против одного	243
6.5. Кривые прецизионность — полнота	248
6.6. Кумулятивный отклик и Lift-кривые	251
6.7. Более сложный способ оценки классификаторов: дубль два	254
6.8. Подведение итогов	266
Глава 7. Оценка регрессоров	270
7.1. Базовые регрессоры	270

7.2. Дополнительные метрики для оценки регрессоров	272
7.3. Графики остатков	282
7.4. Знакомство с понятием стандартизации	290
7.5. Более сложный способ оценки регрессоров: дубль два	295
7.6. Подведение итогов	302

Часть III. Дополнительные методы и основные сведения

Глава 8. Дополнительные методы классификации	309
8.1. Переосмысление процесса классификации	309
8.2. Деревья решений	312
8.3. Классификаторы на основе метода опорных векторов	323
8.4. Логистическая регрессия	335
8.5. Дискриминантный анализ	347
8.6. Допущения, смещения и классификаторы	366
8.7. Сравнение классификаторов: дубль три	369
8.8. Подведение итогов	372
Глава 9. Дополнительные методы регрессии	378
9.1. Линейная регрессия на скамейке штрафников: регуляризация	379
9.2. Регрессия методом опорных векторов	386
9.3. Кусочно-постоянная регрессия	394
9.4. Деревья регрессии	401
9.5. Сравнение регрессоров: дубль три	403
9.6. Подведение итогов	406
Глава 10. Ручное конструирование признаков: манипулирование данными ради удовольствия и прибыли	409
10.1. Конструирование признаков: терминология и мотивация	409
10.2. Отбор признаков и сокращение объема данных: избавление от мусора	414
10.3. Масштабирование признаков	415
10.4. Дискретизация	420
10.5. Кодирование категориальных переменных	423
10.6. Зависимости и взаимодействия	434
10.7. Манипуляции с целевыми переменными	445
10.8. Подведение итогов	451
Глава 11. Настройка гиперпараметров и конвейеры	455
11.1. Модели, параметры, гиперпараметры	456
11.2. Настройка гиперпараметров	458
11.3. Прыжок в рекурсивную кроличью нору: вложенная перекрестная проверка	468

11.4. Конвейеры	477
11.5. Конвейеры и совместная настройка гиперпараметров	480
11.6. Подведение итогов	482

Часть IV. Добавление сложности

Глава 12. Объединение моделей	487
12.1. Ансамбли	487
12.2. Голосующие ансамбли	490
12.3. Бэггинг и случайные леса	491
12.4. Бустинг	500
12.5. Сравнение ансамблевых методов на основе деревьев решений	505
12.6. Подведение итогов	509
Глава 13. Модели, которые конструируют признаки за нас	513
13.1. Отбор признаков	516
13.2. Построение признаков с помощью ядер	537
13.3. Анализ главных компонент: обучение без учителя	557
13.4. Подведение итогов	579
Глава 14. Конструирование признаков для предметных областей: предметно-ориентированное обучение	587
14.1. Работа с текстом	588
14.2. Кластеризация	599
14.3. Работа с изображениями	601
14.4. Подведение итогов	616
Глава 15. Взаимосвязи, расширения и дальнейшие направления	620
15.1. Оптимизация	620
15.2. Построение модели линейной регрессии из примитивных компонентов	624
15.3. Построение модели логистической регрессии из примитивных компонентов	629
15.4. Построение SVM из примитивных компонентов	636
15.5. Нейронные сети	637
15.6. Графовые вероятностные модели	644
15.7. Подведение итогов	654
Приложение А. Листинг модуля mlwpr.py	657
Предметный указатель	665