

УДК 004.4  
ББК 32.972  
Ф70

Флах П.  
Ф70 Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2015. – 400 с.: ил.

**ISBN 978-5-97060-273-7**

Перед вами один из самых интересных учебников по машинному обучению – разделу искусственного интеллекта, изучающего методы построения моделей, способных обучаться, и алгоритмов для их построения и обучения. Автор воздал должное невероятному богатству предмета и не упустил из виду объединяющих принципов. Читатель с первых страниц видит машинное обучение в действии, но без не нужных на первых порах технических деталей. По мере изучения предмета тщательно подобранные примеры, сопровождаемые иллюстрациями, постепенно усложняются.

В книге описан широкий круг логических, геометрических и статистических моделей, затрагиваются и такие находящиеся на переднем крае науки темы, как матричная факторизация и анализ РХП. Особое внимание уделено важнейшей роли признаков. Устоявшаяся терминология дополняется введением в рассмотрение новых полезных концепций. В конце каждой главы приводятся ссылки на дополнительную литературу с авторскими комментариями.

Благодаря всему этому книга задает новый стандарт изучения такой сложной дисциплины как машинное обучение.

Цветные рисунки к книге размещены на нашем сайте [www.dmkpress.com](http://www.dmkpress.com)

УДК 004.4  
ББК 32.972

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press. First published 2012.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-1-107-09639-4 (анг.)  
ISBN 978-5-97060-273-7 (рус.)

© Peter Flach 2012  
© Издание, перевод, ДМК Пресс, 2015

# Содержание

<b>Предисловие .....</b>	<b>11</b>
Как читать эту книгу.....	11
Благодарности.....	12
<b>Пролог: пример машинного обучения.....</b>	<b>14</b>
<b>1 Ингредиенты машинного обучения.....</b>	<b>25</b>
1.1 Задачи: проблемы, решаемые методами машинного обучения .....	25
В поисках структуры .....	27
Оценка качества решения задачи.....	30
1.2 Модели: результат машинного обучения .....	32
Геометрические модели.....	33
Вероятностные модели .....	37
Логические модели.....	44
Группировка и ранжирование .....	49
1.3 Признаки: рабочая лошадка машинного обучения.....	50
Два способа использования признаков.....	52
Отбор и преобразование признаков.....	54
Взаимодействие между признаками .....	56
1.4 Итоги и перспективы.....	59
Что будет в книге дальше.....	61
<b>2 Бинарная классификация и родственные задачи .....</b>	<b>62</b>
2.1 Классификация .....	65
Оценка качества классификации .....	66
Наглядное представление качества классификации.....	70
2.2 Оценивание и ранжирование .....	75
Оценка и визуализация качества ранжирования .....	78

Преобразование ранжировщика в классификатор .....	84
2.3. Оценивание вероятностей классов.....	87
Качество оценивания вероятностей классов.....	88
Преобразование ранжировщиков в оценки вероятностей классов.....	91
2.4 Бинарная классификация и родственные задачи: итоги и дополнительная литература .....	93
<b>3 За пределами бинарной классификации .....</b>	<b>96</b>
3.1 Когда классов больше двух .....	96
Многоклассовая классификация .....	96
Многоклассовые оценки и вероятности .....	101
3.2 Регрессия .....	105
3.3 Обучение без учителя и дескриптивные модели.....	108
Прогностическая и дескриптивная кластеризация .....	109
Другие дескриптивные модели .....	114
3.4 За пределами бинарной классификации: итоги и литература для дальнейшего чтения .....	116
<b>4 Концептуальное обучение .....</b>	<b>118</b>
4.1 Пространство гипотез .....	119
Наименьшее обобщение .....	120
Внутренняя дизъюнкция.....	122
4.2 Пути в пространстве гипотез .....	124
Наиболее общие непротиворечивые гипотезы.....	128
Замкнутые концепты.....	130
4.3 За пределами конъюнктивных концептов .....	130
Применение логики первого порядка.....	135
4.4 Обучаемость .....	136
4.5 Концептуальное обучение: итоги и литература для дальнейшего чтения.....	139
<b>5 Древоподобные модели.....</b>	<b>142</b>
5.1 Решающие деревья .....	146
5.2 Деревья ранжирования и оценивания вероятностей.....	151
Чувствительность к асимметричному распределению по классам.....	156
5.3 Обучение деревьев как уменьшение дисперсии .....	161
Деревья регрессии .....	161
Кластеризующие деревья.....	165

5.4	Древовидные модели: итоги и литература для дальнейшего чтения .....	168
<b>6</b>	<b>Модели на основе правил.....</b>	<b>170</b>
6.1	Обучение упорядоченных списков правил.....	170
	Списки правил для ранжирования и оценивания вероятностей.....	176
6.2	Обучение неупорядоченных множеств правил.....	179
	Применение множеств правил для ранжирования и оценивания вероятностей.....	183
	Более пристальный взгляд на перекрытие правил .....	187
6.3	Обучение дескриптивных моделей на основе правил .....	189
	Обучение правил для выявления подгрупп .....	190
	Добыча ассоциативных правил.....	194
6.4	Обучение правил первого порядка.....	199
6.5	Модели на основе правил: итоги и литература для дальнейшего чтения.....	203
<b>7</b>	<b>Линейные модели.....</b>	<b>206</b>
7.1	Метод наименьших квадратов .....	208
	Многомерная линейная регрессия .....	212
	Регуляризованная регрессия .....	216
	Применение регрессии по методу наименьших квадратов к задаче классификации .....	217
7.2	Перцептрон .....	218
7.3	Метод опорных векторов .....	223
	Метод опорных векторов с мягким зазором.....	228
7.4	Получение вероятностей от линейных классификаторов.....	231
7.5	За пределами линейности – ядерные методы.....	236
7.6	Линейные модели: итоги и литература для дальнейшего чтения.....	239
<b>8</b>	<b>Метрические модели .....</b>	<b>242</b>
8.1	Так много дорог.....	242
8.2	Соседи и эталоны .....	248
8.3	Классификация по ближайшему соседу.....	253
8.4	Метрическая кластеризация.....	256
	Алгоритм К средних .....	259
	Кластеризация вокруг медоидов .....	261
	Силуэты.....	262
8.5	Иерархическая кластеризация.....	264

8.6	От ядер к расстояниям.....	269
8.7	Метрические модели: итоги и литература для дальнейшего чтения.....	270
<b>9</b>	<b>Вероятностные модели.....</b>	<b>273</b>
9.1	Нормальное распределение и его геометрические интерпретации.....	277
9.2	Вероятностные модели для категориальных данных .....	284
	Использование наивной байесовской модели для классификации.....	286
	Обучение наивной байесовской модели.....	289
9.3	Дискриминантное обучение путем оптимизации условного правдоподобия .....	293
9.4	Вероятностные модели со скрытыми переменными .....	297
	ЕМ-алгоритм .....	299
	Гауссовы смесовые модели .....	300
9.5	Модели на основе сжатия .....	304
9.6	Вероятностные модели: итоги и литература для дальнейшего чтения.....	306
<b>10</b>	<b>Признаки.....</b>	<b>310</b>
10.1	Виды признаков .....	310
	Вычисления с признаками.....	311
	Категориальные, порядковые и количественные признаки .....	315
	Структурированные признаки.....	317
10.2	Преобразования признаков.....	318
	Задание порога и дискретизация.....	319
	Нормировка и калибровка.....	325
	Неполные признаки .....	333
10.3	Конструирование и отбор признаков .....	334
	Преобразование и разложение матриц .....	336
10.4	Признак: итоги и литература для дальнейшего чтения .....	339
<b>11</b>	<b>Ансамбли моделей .....</b>	<b>342</b>
11.1	Баггинг и случайные леса .....	343
11.2	Усиление.....	345
	Обучение усиленных правил.....	349
11.3	Карта ансамблевого ландшафта .....	350
	Смещение, дисперсия и зазоры.....	350
	Другие ансамблевые методы.....	352
	Метаобучение.....	352

11.4 Ансамбли моделей: итоги и литература для дальнейшего чтения .....	353
<b>12 Эксперименты в машинном обучении .....</b>	<b>355</b>
12.1 Что измерять .....	356
12.2 Как измерять .....	360
12.3 Как интерпретировать.....	362
Интерпретация результатов, полученных на нескольких наборах данных.....	365
12.4 Эксперименты в машинном обучении: итоги и литература для дальнейшего чтения .....	368
<b>Эпилог: что дальше? .....</b>	<b>371</b>
<b>Что нужно запомнить.....</b>	<b>373</b>
<b>Библиография.....</b>	<b>376</b>
<b>Предметный указатель.....</b>	<b>387</b>