

Оглавление

Об авторе	20
От научного редактора перевода.....	20
От издательства	22
ПРЕАМБУЛА	23
Глава 1. Финансовое машинное обучение как отдельная дисциплина	24
1.1. Актуальность	24
1.2. Основная причина безуспешности проектов финансового машинного обучения	25
1.2.1. Сизифова парадигма.....	26
1.2.2. Метастратегическая парадигма.....	27
1.3. Структура книги.....	27
1.3.1. Структура в виде производственной цепочки.....	28
1.3.2. Структура по компонентам стратегии.....	32
1.3.3. Структура по распространенной ошибке	36
1.4. Целевая аудитория	37
1.5. Справочная информация	38
1.6. Часто задаваемые вопросы	39
1.7. Благодарности	44
Упражнения.....	45
ЧАСТЬ 1. АНАЛИЗ ДАННЫХ	47
Глава 2. Структуры финансовых данных	48
2.1. Актуальность	48
2.2. Основные типы финансовых данных	48
2.2.1. Базовые данные.....	49
2.2.2. Данные рынка.....	50
2.2.3. Аналитические данные	50

2.2.4. Альтернативные данные	50
2.3. Бары	51
2.3.1. Стандартные бары	51
2.3.2. Информационные гистограммы	55
2.4. Работа с мультипродуктовыми рядами	60
2.4.1. Трюк ETF	60
2.4.2. Веса метода главных компонент (МГК)	63
2.4.3. Перенесение одного фьючерсного контракта	65
2.5. Отбор признаков	67
2.5.1. Отбор с целью сокращения	67
2.5.2. Событийно-управляемый отбор	68
Упражнения	70
Глава 3. Маркировка	72
3.1. Актуальность	72
3.2. Метод фиксированного временного горизонта	72
3.3. Вычисление динамических порогов	74
3.4. Тройной барьерный метод	74
3.5. Выяснение стороны и размера ставки	78
3.6. Метамаркировка	80
3.7. Как использовать метамаркировку	82
3.8. Квантоментальный способ	84
3.9. Исключение ненужных меток	85
Упражнения	86
Глава 4. Веса выборки	88
4.1. Актуальность	88
4.2. Накладывающиеся исходы	88
4.3. Число одновременных меток	89
4.4. Средняя уникальность метки	90
4.5. Бэггинг классификаторов и уникальности	91
4.5.1. Последовательное бутстрапирование	93
4.5.2. Реализация последовательного бутстрапирования	94
4.5.3. Числовой пример	95
4.5.4. Эксперименты Монте-Карло	96
4.6. Атрибутирование финансовых возвратов	98
4.7. Временной спад, или эрозия	99
4.8. Веса классов	101
Упражнения	102

Глава 5. Дробно-дифференцированные признаки	104
5.1. Актуальность	104
5.2. Дилемма «стационарность или память»	104
5.3. Обзор публикаций	105
5.4. Метод	106
5.4.1. Долгая память	107
5.4.2. Итеративное оценивание	107
5.4.3. Сходимость	109
5.5. Применение	110
5.5.1. Расширяющееся окно	110
5.5.2. Дробное дифференцирование с окном фиксированной ширины	112
5.6. Стационарность с максимальным сохранением памяти	114
5.7. Заключение	116
Упражнения	119
ЧАСТЬ 2. МОДЕЛИРОВАНИЕ	121
Глава 6. Ансамблевые методы	122
6.1. Актуальность	122
6.2. Три источника ошибок	122
6.3. Агрегация бутстрапов	123
6.3.1. Сокращение дисперсии	124
6.3.2. Улучшенная точность	125
6.3.3. Избыточность наблюдений	127
6.4. Случайный лес	127
6.5. Бустирование	129
6.6. Бэггинг vs бустинг в финансах	130
6.7. Бэггинг для масштабируемости	131
Упражнения	132
Глава 7. Перекрестная проверка в финансах	133
7.1. Актуальность	133
7.2. Цель перекрестной проверки	133
7.3. Почему перекрестная проверка по k блокам оказывается безуспешной в финансах	135
7.4. Решение: прочищенная k -блочная перекрестная проверка	136
7.4.1. Очищение набора данных для обучения	136
7.4.2. Эмбарго	138
7.4.3. Класс прочищенной k -блочной перекрестной проверки	139

7.5. Дефекты реализации перекрестной проверки в библиотеке sklearn	140
Упражнения.....	141
Глава 8. Важность признаков	143
8.1. Актуальность	143
8.2. Значимость признаков	143
8.3. Важность признаков и эффекты замещения.....	144
8.3.1. Среднее снижение в примесности.....	145
8.3.2. Среднее снижение точности	146
8.4. Важность признаков без эффектов замещения	148
8.4.1. Однопризнаковая важность	148
8.4.2. Ортогональные признаки.....	149
8.5. Параллелизованная важность признаков против стековой.....	152
8.6. Эксперименты с синтетическими данными	153
Упражнения.....	159
Глава 9. Регулировка гиперпараметров с помощью перекрестной проверки.....	160
9.1. Актуальность	160
9.2. Перекрестная проверка с помощью решеточного поиска	160
9.3. Перекрестная проверка с помощью рандомизированного поиска.....	162
9.3.1. Логарифмически равномерное распределение.....	163
9.4. Балльное оценивание и регулировка гиперпараметров.....	165
Упражнения.....	167
ЧАСТЬ 3. БЭКТЕСТИРОВАНИЕ	169
Глава 10. Выставление размера ставки	170
10.1. Актуальность	170
10.2. Независимые от стратегии подходы к выставлению размеров.....	170
10.3. Выставление размера из предсказанных вероятностей	172
10.4. Усреднение активных ставок.....	173
10.5. Дискретизация размера	174
10.6. Динамические размеры ставок и лимитные цены	175
Упражнения.....	178
Глава 11. Опасности бэкестирования.....	180
11.1. Актуальность	180
11.2. Миссия невыполнима: безупречный бэкест.....	180
11.3. Даже если ваш бэкест безупречен, он, вероятнее всего, будет ошибочен	182
11.4. Бэкестирование — это не исследовательский инструмент.....	182

11.5. Несколько общих рекомендаций	183
11.6. Выбор стратегии	185
Упражнения.....	189
Глава 12. Бэктестирование через кросс-валидацию	190
12.1. Актуальность	190
12.2. Прямой метод	190
12.2.1. Ловушки прямого метода	191
12.3. Перекрестно-проверочный метод.....	192
12.4. Комбинаторный прочищенный перекрестно-проверочный метод	193
12.4.1. Комбинаторное дробление на подразделы	194
12.4.2. Алгоритм бэктестирования на основе комбинаторной прочищенной перекрестной проверки	195
12.4.3. Примеры	196
12.5. Как комбинаторная прочищенная перекрестная проверка справляется с бэктестовой переподгонкой	196
Упражнения.....	198
Глава 13. Бэктестирование на синтетических данных	200
13.1. Актуальность	200
13.2. Правила трейдинга	200
13.3. Проблема.....	201
13.4. Наш математический каркас	203
13.5. Численное определение оптимальных торговых правил.....	204
13.5.1. Алгоритм.....	204
13.5.2. Реализация	206
13.6. Экспериментальные результаты.....	207
13.6.1. Случаи с нулевым долгосрочным равновесием.....	209
13.6.2. Случаи с положительным долгосрочным равновесием	213
13.6.3. Случаи с отрицательным долгосрочным равновесием	216
13.7. Выводы	225
Упражнения.....	225
14. Статистические показатели бэктеста.....	227
14.1. Актуальность.....	227
14.2. Виды статистических показателей бэктеста	227
14.3. Основные характеристики.....	228
14.4. Результативность	230
14.4.1. Взвешенная по времени возвратность.....	231
14.5. Интервалы	232

14.5.1. Концентрация финансовых возвратов	232
14.5.2. Просадка и время нахождения ниже уровня воды	234
14.5.3. Статистические показатели интервалов для оценивания результативности	235
14.6. Дефицит реализации	235
14.7. Эффективность	236
14.7.1. Коэффициент Шарпа	236
14.7.2. Вероятностный коэффициент Шарпа	236
14.7.3. Дефлированный коэффициент Шарпа	238
14.7.4. Статистические показатели эффективности	239
14.8. Классификационные балльные оценки	240
14.9. Атрибутирование	242
Упражнения	243
Глава 15. Понимание риска стратегии	245
15.1. Актуальность	245
15.2. Симметричные выплаты	245
15.3. Асимметричные выплаты	247
15.4. Вероятность неуспешности стратегии	250
15.4.1. Алгоритм	251
15.4.2. Реализация	252
Упражнения	253
Глава 16. Распределение финансовых активов	255
16.1. Актуальность	255
16.2. Проблема выпуклой портфельной оптимизации	255
16.3. Проклятие Марковица	256
16.4. От геометрических связей к иерархическим	258
16.4.1. Деревовидная кластеризация	259
16.4.2. Квазидиагонализация	263
16.4.3. Рекурсивное дробление пополам	264
16.5. Численный пример	266
16.6. Вневыборочные симуляции Монте-Карло	269
16.7. Дальнейшие исследования	272
16.8. Заключение	274
Дополнение	275
16.A.1. Корреляционный метрический показатель	275
16.A.2. Инверсно-дисперсное размещение	276
16.A.3. Воспроизведение численного примера	277

16.А.4. Воспроизведение эксперимента Монте-Карло	279
Упражнения.....	281

ЧАСТЬ 4. ПОЛЕЗНЫЕ ФИНАНСОВЫЕ ПРИЗНАКИ 283

Глава 17. Структурные сдвиги.....	284
17.1. Актуальность.....	284
17.2. Типы проверок на структурные сдвиги.....	284
17.3. Проверки на основе фильтра CUSUM.....	285
17.3.1. Проверка CUSUM Брауна—Дарбина—Эванса на рекурсивных остатках.....	285
17.3.2. Проверка CUSUM Чу—Стинчкомба—Уайта на уровнях.....	286
17.4. Проверки взрываемости.....	286
17.4.1. Проверка Дики—Фуллера по типу Чоу.....	287
17.4.2. Супремально расширенный тест Дики—Фуллера.....	288
17.4.3. Суб- и супермартингейловые проверки.....	296
Упражнения.....	297
Глава 18. Энтропийные признаки.....	299
18.1. Актуальность.....	299
18.2. Энтропия Шеннона.....	299
18.3. Подстановочный (или максимально правдоподобный) оценщик.....	301
18.4. Оценщики на основе алгоритма LZ.....	302
18.5. Схемы кодирования.....	306
18.5.1. Двоичное кодирование.....	306
18.5.2. Квантильное кодирование.....	306
18.5.3. Сигма-кодирование.....	307
18.6. Энтропия гауссова процесса.....	307
18.7. Энтропия и обобщенное среднее.....	310
18.8. Несколько финансовых приложений энтропии.....	312
18.8.1. Рыночная эффективность.....	312
18.8.2. Генерирование максимальной энтропии.....	312
18.8.3. Концентрация портфеля.....	312
18.8.4. Микроструктура рынка.....	313
Упражнения.....	315
Глава 19. Микроструктурные признаки.....	317
19.1. Актуальность.....	317
19.2. Обзор литературы.....	317
19.3. Первое поколение: ценовые последовательности.....	318

19.3.1. Тиковое правило.....	318
19.3.2. Модель Ролла	319
19.3.3. Оценщик волатильности максимум-минимум.....	320
19.3.4. Корвин и Шульц.....	321
19.4. Второе поколение: стратегические модели сделок.....	323
19.4.1. Лямбда Кайла	324
19.4.2. Лямбда Амихуда.....	326
19.4.3. Лямбда Хасбрука.....	326
19.5. Третье поколение: модели последовательных сделок	327
19.5.1. Вероятность информационно обусловленной торговли	328
19.5.2. Объемно-синхронизированная вероятность информированной торговли	329
19.6. Дополнительные признаки из микроструктурных совокупностей данных	330
19.6.1. Распределение объемов ордеров	331
19.6.2. Скорости отмены, лимитные и рыночные ордера	331
19.6.3. Исполнительные алгоритмы TWAP	332
19.6.4. Опционные рынки.....	333
19.6.5. Внутрисядовая корреляция ориентированного (по знаку) потока ордеров.....	334
19.7. Что такое микроструктурная информация?.....	334
Упражнения.....	336

ЧАСТЬ 5. РЕЦЕПТЫ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ 339

Глава 20. Мультиобработка и векторизация.....	340
20.1. Актуальность	340
20.2. Пример векторизации	340
20.3. Однопоточность против многопоточности и мультиобработки.....	341
20.4. Атомы и молекулы	343
20.4.1. Линейные подразделы	343
20.4.2. Подразделения с дважды вложенными циклами	344
20.5. Мультиобрабатывающие механизмы	346
20.5.1. Подготовка заданий	347
20.5.2. Асинхронные вызовы	349
20.5.3. Разворачивание функции обратного вызова	349
20.5.4. Консервация/расконсервация объектов.....	350
20.5.5. Сокращение результата	350

20.6. Пример мультиобработки.....	352
Упражнения.....	353
Глава 21. Метод полного перебора и квантовые компьютеры	355
21.1. Актуальность	355
21.2. Комбинаторная оптимизация	355
21.3. Целевая функция.....	356
21.4. Задача	357
21.5. Целочисленно-оптимизационный подход	357
21.5.1. Подразделения по методу голубиных клеток.....	358
21.5.2. Допустимые статические решения	359
21.5.3. Оценивание траекторий.....	360
21.6. Численный пример.....	361
21.6.1. Случайные матрицы.....	361
21.6.2. Статическое решение.....	362
21.6.3. Динамическое решение.....	363
Упражнения.....	363
Глава 22. Технологии высокопроизводительного вычислительного интеллекта и прогнозирования	365
22.1. Актуальность	365
22.2. Регулятивная реакция на молниеносный обвал 2010 года.....	366
22.3. История вопроса	366
22.4. Аппаратное обеспечение для высокопроизводительных вычислений	368
22.5. Программное обеспечение высокопроизводительных вычислений	372
22.5.1. Интерфейс передачи сообщений.....	373
22.5.2. Иерархический формат данных 5 (HDF5)	373
22.5.3. Обработка прямо на месте	374
22.5.4. Конвергенция.....	375
22.6. Примеры использования	375
22.6.1. Поиск сверхновой звезды.....	376
22.6.2. Пятна в термоядерной плазме	377
22.6.3. Внутрисуточное пиковое потребление электроэнергии	379
22.6.4. Молниеносный обвал 2010 года	384
22.6.5. Калибровка объемно-синхронизированной вероятности информированной торговли	386
22.6.6. Выявление высокочастотных событий с помощью неравномерного быстрого преобразования Фурье	388
22.7. Итоги и призыв к сотрудничеству.....	389
22.8. Благодарности	391

ПРИЛОЖЕНИЯ 393**Приложение А.** О книге «Машинное обучение: алгоритмы для бизнеса»

Лопеза де Прадо.....	394
А.1. Структуры данных	394
А.2. Статистические свойства и преобразования стационарности	396
А.3. Маркировка для самообучения	397
А.3.1. Тройной барьерный метод.....	397
А.3.2. Метамаркировка	398
А.4. Обучающиеся алгоритмы для направления и размера ставки	399
А.4.1. Агрегирование бутстраповских выборок (бэггированные классификаторы).....	399
А.4.2. Случайные леса	400
А.4.3. Важность признаков	400
А.4.4. Перекрестная проверка	401
А.5. Дальнейшие исследования	402
Приложение Б. Глоссарий.....	403
Приложение В. Справочные материалы и библиография	413