

Машинное обучение

ПРОЕКТИРОВАНИЕ СИСТЕМ
ОТ ИДЕИ ДО РЕАЛИЗАЦИИ

ВАЛЕРИЙ БАБУШКИН, АРСЕНИЙ КРАВЧЕНКО



Санкт-Петербург · Москва · Минск

2026

Краткое содержание

Часть 1. Подготовка к работе

Глава 1. Основы проектирования систем машинного обучения.....	24
Глава 2. В чем проблема?	40
Глава 3. Предварительное изучение.....	57
Глава 4. Дизайн-документ	77

Часть 2. Первые шаги

Глава 5. Функции потерь и метрики	96
Глава 6. Подготовка датасета.....	125
Глава 7. Схемы валидации.....	154
Глава 8. Базовая модель	178

Часть 3. Промежуточные этапы

Глава 9. Анализ ошибок.....	198
Глава 10. Пайплайны обучения	233
Глава 11. Признаки и генерация признаков.....	254
Глава 12. Измерение и представление результатов	290

Часть 4. Интеграция и масштабирование

Глава 13. Интеграция	322
Глава 14. Мониторинг и надежность	343
Глава 15. Сервинг и оптимизация инференса	377
Глава 16. Владение и сопровождение	399

Оглавление

От издательства.....	14
О научном редакторе русского издания	14
Предисловие	15
Благодарности	16
О книге	18
Для кого эта книга?	18
Структура книги	19
Форум liveBook	20
Об авторах	21
Иллюстрация на обложке.....	22

Часть 1. Подготовка к работе

Глава 1. Основы проектирования систем машинного обучения.....	24
1.1. Что такое проектирование ML-систем?	25
1.1.1. Почему проектирование ML-систем так важно	30
1.1.2. Истоки проектирования ML-систем	30
1.2. Структура книги.....	32
1.3. Где могут быть полезны принципы проектирования ML-систем	35
Итоги	39
Глава 2. В чем проблема?	40
2.1. Пространство проблем и пространство решений	41
2.2. Поиск проблемы	45
2.2.1. Как прийти к решению проблемы с помощью ML-системы	49

2.3. Риски, ограничения и возможные последствия	50
2.4. Цена ошибки	53
Итоги	55
Глава 3. Предварительное изучение	57
3.1. Какие проблемы способны вдохновлять?	58
3.2. Создать или купить — технологии с открытым исходным кодом или коммерческие	60
3.2.1. Создать или купить	60
3.2.2. Технологии с открытым исходным кодом или коммерческие решения	63
3.3. Декомпозиция проблемы	63
3.4. Выбор подходящего уровня инновации	68
3.4.1. Как найти эффективное решение?	70
3.4.2. Работа в пространстве решений: пример из практики	72
Итоги	75
Глава 4. Дизайн-документ	77
4.1. Распространенные мифы о дизайн-документе	78
4.1.1. Миф № 1. Дизайн-документы полезны в крупных компаниях, но не в стартапах	78
4.1.2. Миф № 2. Дизайн-документы нужны только в сложных проектах	79
4.1.3. Миф № 3. Дизайн-документ пишется по шаблону	79
4.1.4. Миф № 4. Каждый дизайн-документ должен описывать развернутую систему	80
4.2. Цели и антицели	81
4.3. Структура дизайн-документа	83
4.4. Ревью дизайн-документа	88
4.4.1. Пример ревью дизайн-документа	90
4.5. Дизайн-документ — это живой организм	92
Итоги	94

Часть 2. Первые шаги

Глава 5. Функции потерь и метрики	96
5.1. Функции потерь	97
5.1.1. Выбор функции потерь для моделей глубокого обучения	100

5.2. Метрики.....	103
5.2.1. Метрики согласованности	112
5.2.2. Офлайн- и онлайн-метрики, прокси-метрики и иерархия метрик	114
5.3. Дизайн-документ: добавляем функции потерь и метрики.....	117
5.3.1. Метрики и функции потерь для Supermegaretail.....	118
5.3.2. Метрики и функции потерь для PhotoStock Inc.....	122
5.3.3. Вывод.....	124
Итоги.....	124
Глава 6. Подготовка датасета.....	125
6.1. Источники данных.....	126
6.2. Подготовка датасета	129
6.2.1. ETL.....	129
6.2.2. Фильтрация.....	130
6.2.3. Генерация признаков.....	131
6.2.4. Разметка	131
6.3. Данные и метаданные	136
6.4. Сколько нужно данных?.....	138
6.5. Проблема курицы и яйца.....	141
6.6. Критерий устойчивого пайплайна данных	143
6.7. Дизайн-документ: датасет	146
6.7.1. Датасет для Supermegaretail	147
6.7.2. Датасет для PhotoStock Inc.	150
Итоги.....	152
Глава 7. Схемы валидации	154
7.1. Достоверная оценка.....	155
7.2. Стандартные схемы.....	156
7.2.1. Отложенная выборка.....	156
7.2.2. Кросс-валидация	157
7.2.3. Выбор значения K.....	158
7.2.4. Валидация временных рядов	159
7.3. Нестандартные схемы.....	162
7.3.1. Вложенная валидация	163
7.3.2. Адверсариальная валидация	164
7.3.3. Оценка использования утечек данных	165

7.4. Процедура раздельного обновления	166
7.5. Дизайн-документ: выбор схем валидации.....	173
7.5.1. Схемы валидации для Supermegaretail.....	173
7.5.2. Схемы валидации для PhotoStock Inc.....	176
Итоги.....	177
Глава 8. Базовая модель	178
8.1. Что такое базовая модель.....	179
8.2. Константные базовые модели.....	182
8.2.1. Зачем нужны константные базовые модели?.....	184
8.3. Базовые модели и базовые признаки	185
8.4. Базовые модели глубокого обучения.....	188
8.5. Сравнение базовых моделей	189
8.6. Дизайн-документ: базовые модели.....	192
8.6.1. Базовые модели для Supermegaretail	192
8.6.2. Базовые модели для компании PhotoStock Inc.....	194
Итоги.....	195
 Часть 3. Промежуточные этапы	
Глава 9. Анализ ошибок	198
9.1. Анализ кривой обучения	199
9.1.1. Переобучение и недообучение.....	200
9.1.2. Кривая потерь.....	201
9.1.3. Анализ кривых потерь.....	202
9.1.4. Кривая обучения по модели.....	206
9.1.5. Кривая обучения по объему выборки.....	207
9.1.6. Двойной спуск.....	207
9.2. Анализ остатков.....	209
9.2.1. Цели анализа остатков	210
9.2.2. Предположения о модели	211
9.2.3. Распределение остатков.....	215
9.2.4. Справедливость остатков.....	217
9.2.5. Недооценка и переоценка.....	219
9.2.6. Кривые эластичности	219
9.3. Поиск закономерностей в остатках	222
9.3.1. Анализ наихудших/наилучших случаев.....	222
9.3.2. Адверсариальная валидация.....	223

9.3.3. Разнообразие группового анализа	224
9.3.4. Анализ крайних случаев	225
9.4. Дизайн-документ: анализ ошибок	226
9.4.1. Анализ ошибок для Supermegaretail	226
9.4.2. Анализ ошибок для компании PhotoStock Inc.	230
Итоги	232
Глава 10. Пайплайны обучения	233
10.1. Что такое пайплайн обучения?	233
10.1.1. Пайплайн обучения vs пайплайн инференса	235
10.2. Инструменты и платформы	238
10.3. Масштабируемость	241
10.4. Конфигурируемость	242
10.5. Тестирование	245
10.5.1. Тестирование на основе свойств	247
10.6. Дизайн-документ: пайплайны обучения	249
10.6.1. Пайплайн обучения для Supermegaretail	249
10.6.2. Пайплайн обучения для PhotoStock Inc	251
Итоги	252
Глава 11. Признаки и генерация признаков	254
11.1. Что такое генерация признаков?	255
11.1.1. Критерии хороших и плохих признаков	256
11.1.2. Основы генерации признаков	258
11.1.3. Прогнозы модели как признак	260
11.2. Анализ важности признаков	261
11.2.1. Классификация методов	263
11.2.2. Компромисс между точностью и интерпретируемостью	265
11.2.3. Важность признаков в глубоком обучении	266
11.3. Отбор признаков	269
11.3.1. Генерация признаков vs отбор признаков	269
11.3.2. Цели и возможные недостатки	270
11.3.3. Методы отбора признаков	272
11.4. Хранилище признаков	276
11.4.1. Хранилище признаков: плюсы и минусы	277
11.4.2. Желательные свойства хранилища признаков	280
11.4.3. Каталог признаков	284

11.5. Дизайн-документ: генерация признаков	284
11.5.1. Признаки для Supermegaretail.....	284
11.5.2. Признаки для PhotoStock Inc.....	287
Итоги.....	289
Глава 12. Измерение и представление результатов	290
12.1. Измерение результатов.....	291
12.1.1. Эффективность модели	291
12.1.2. Переход к бизнес-метрикам	292
12.1.3. Среда моделирования.....	293
12.1.4. Оценка человеком.....	298
12.2. А/В-тестирование.....	298
12.2.1. Планирование эксперимента.....	299
12.2.2. Стратегия разделения	301
12.2.3. Выбор метрик.....	303
12.2.4. Статистические критерии.....	303
12.2.5. Моделирование экспериментов.....	304
12.2.6. Когда проведение А/В-тестирования невозможно.....	306
12.3. Представление результатов.....	306
12.3.1. Контрольные и вспомогательные метрики.....	307
12.3.2. Мониторинг прироста.....	307
12.3.3. Когда завершать эксперимент.....	308
12.3.4. Что указывать в отчете.....	309
12.3.5. Отчетный документ	310
12.4. Дизайн-документ: измерение и отчетность	311
12.4.1. Измерение и отчетность для Supermegaretail.....	311
12.4.2. Измерение и отчетность для компании PhotoStock Inc.	313
Итоги.....	320

Часть 4. Интеграция и масштабирование

Глава 13. Интеграция	322
13.1. Проектирование API.....	323
13.1.1. Работа с API.....	327
13.2. Релизный цикл	329
13.3. Эксплуатация системы	333
13.3.1. Технические связи	333
13.3.2. Нетехнические связи	334

13.4. Переопределения и фоллбэк-решения.....	334
13.5. Дизайн-документ: интеграция	336
13.5.1. Интеграция для Supermegaretail	336
13.5.2. Интеграция для PhotoStock Inc.....	339
Итоги.....	342
Глава 14. Мониторинг и надежность	343
14.1. Зачем нужен мониторинг.....	344
14.1.1. Входные данные	345
14.1.2. Модель.....	345
14.1.3. Выход модели.....	346
14.1.4. Постобработка / Принятие решений.....	347
14.2. Работоспособность программного обеспечения	348
14.3. Качество и целостность данных.....	350
14.3.1. Проблемы обработки.....	350
14.3.2. Повреждение источника данных.....	351
14.3.3. Каскадные/вышестоящие модели	352
14.3.4. Изменение схемы данных.....	353
14.3.5. Перекос между обучением и эксплуатацией	353
14.3.6. Как осуществлять мониторинг и реагировать на проблемы	354
14.4. Качество и релевантность модели	358
14.4.1. Дрейф данных	361
14.4.2. Дрейф концепции	361
14.4.3. Как осуществлять мониторинг	363
14.4.4. Как реагировать.....	365
14.5. Дизайн-документ: мониторинг	371
14.5.1. Мониторинг для Supermegaretail	371
14.5.2. Мониторинг для PhotoStock Inc.	374
Итоги.....	376
Глава 15. Сервинг и оптимизация инференса	377
15.1. Сервинг и инференс: сложности	378
15.2. Компромиссы и паттерны.....	381
15.2.1. Компромиссы	381
15.2.2. Паттерны.....	383

15.3. Инструменты и фреймворки	385
15.3.1. Выбор фреймворка.....	385
15.3.2. Бессерверный инференс.....	388
15.4. Оптимизация пайплайнов инференса.....	390
15.4.1. Начинаем с профилирования.....	391
15.4.2. Минимальная оптимизация — наилучшая оптимизация.....	393
15.5. Дизайн-документ: сервинг и инференс.....	394
15.5.1. Сервинг и инференс для Supermegaretail.....	394
15.5.2. Сервинг и инференс для PhotoStock Inc.....	395
Итоги.....	398
Глава 16. Владение и сопровождение	399
16.1. Ответственность.....	400
16.2. Фактор автобуса.....	405
16.2.1. Чем опасна чрезмерная эффективность?	406
16.2.2. Чем опасна чрезмерная избыточность?	407
16.2.3. Когда и как учитывать фактор автобуса.....	407
16.3. Документация	408
16.4. Сложность	410
16.5. Сопровождение и владение: Supermegaretail и PhotoStock Inc.....	414
Итоги.....	415