

# Оглавление

От издательства .....	14
О научном редакторе русского издания .....	14
<b>Математика в машинном обучении</b> .....	<b>15</b>
<b>Условные обозначения</b> .....	<b>17</b>
Список аббревиатур и сокращений .....	19
<b>Предисловие</b> .....	<b>20</b>
Зачем нужна еще одна книга по машинному обучению? .....	21
Какова целевая аудитория книги? .....	22
<b>Благодарности</b> .....	<b>24</b>

## ЧАСТЬ I

### Математические основы

<b>Глава 1. Введение и мотивация</b> .....	<b>28</b>
1.1. Поиск интуитивно понятных формулировок .....	29
1.2. Два способа читать эту книгу .....	30
1.3. Упражнения и обратная связь .....	34
<b>Глава 2. Линейная алгебра</b> .....	<b>35</b>
2.1. Системы линейных уравнений .....	38
2.2. Матрицы .....	41
2.2.1. Сложение и перемножение матриц .....	42
2.2.2. Обращение и транспонирование .....	44
2.2.3. Умножение на скаляр .....	46
2.2.4. Компактное представление системы уравнений .....	47

2.3. Решение систем линейных уравнений . . . . .	47
2.3.1. Частное и общее решение . . . . .	48
2.3.2. Элементарные преобразования . . . . .	49
2.3.3. Прием с $-1$ . . . . .	54
2.3.4. Алгоритмы для решения системы линейных уравнений . . . . .	57
2.4. Векторные пространства . . . . .	58
2.4.1. Группы . . . . .	58
2.4.2. Векторные пространства . . . . .	60
2.4.3. Векторные подпространства . . . . .	62
2.5. Линейная независимость . . . . .	63
2.6. Базис и ранг . . . . .	68
2.6.1. Генерация множества и базиса . . . . .	69
2.6.2. Ранг . . . . .	72
2.7. Линейные отображения . . . . .	73
2.7.1. Матричное представление линейных отображений . . . . .	75
2.7.2. Изменение базиса . . . . .	79
2.7.3. Образ и ядро . . . . .	85
2.8. Аффинные пространства . . . . .	88
2.8.1. Аффинные подпространства . . . . .	89
2.8.2. Аффинные отображения . . . . .	90
2.9. Дополнительное чтение . . . . .	91
Упражнения . . . . .	92
<b>Глава 3. Аналитическая геометрия . . . . .</b>	<b>100</b>
3.1. Нормы . . . . .	100
3.2. Внутренние произведения . . . . .	102
3.2.1. Скалярное произведение . . . . .	103
3.2.2. Общие внутренние произведения . . . . .	103
3.2.3. Симметричные положительно определенные матрицы . . . . .	104
3.3. Длины и расстояния . . . . .	106
3.4. Углы и ортогональность . . . . .	108
3.5. Ортонормированный базис . . . . .	111
3.6. Ортогональное дополнение . . . . .	112
3.7. Внутреннее произведение функций . . . . .	113

3.8. Ортогональные проекции . . . . .	114
3.8.1. Проекция на одномерные подпространства (прямые) . . . . .	116
3.8.2. Проекция на общие подпространства . . . . .	119
3.8.3. Ортогонализация Грама — Шмидта . . . . .	124
3.8.4. Проекция на аффинные подпространства . . . . .	125
3.9. Повороты . . . . .	126
3.9.1. Повороты в $\mathbb{R}^2$ . . . . .	127
3.9.2. Повороты в $\mathbb{R}^3$ . . . . .	128
3.9.3. Поворот в $n$ измерениях . . . . .	129
3.9.4. Свойства поворотов . . . . .	130
3.10. Дополнительное чтение . . . . .	130
Упражнения . . . . .	131
<b>Глава 4. Матричные разложения . . . . .</b>	<b>134</b>
4.1. Детерминант и след . . . . .	136
4.2. Собственные значения и собственные векторы . . . . .	142
4.2.1. Графическая интуиция в двух измерениях . . . . .	147
4.3. Разложение Холецкого . . . . .	154
4.4. Собственное разложение и диагонализация . . . . .	156
4.4.1. Геометрическая интуиция для собственного разложения . . . . .	158
4.5. Разложение по сингулярным значениям . . . . .	160
4.5.1. Геометрические интуиции для SVD . . . . .	161
4.5.2. Построение SVD . . . . .	164
4.5.3. Разложение на собственные значения и разложение на сингулярные значения . . . . .	169
4.6. Матричное приближение . . . . .	173
4.7. Матричная филогения . . . . .	178
4.8. Дополнительное чтение . . . . .	180
Упражнения . . . . .	182
<b>Глава 5. Векторный анализ . . . . .</b>	<b>185</b>
5.1. Дифференцирование функций одной переменной . . . . .	187
5.1.1. Ряд Тейлора . . . . .	189
5.1.2. Правила дифференцирования . . . . .	192

5.2. Частные производные и градиенты . . . . .	193
5.2.1. Основные правила взятия частных производных . . . . .	195
5.2.2. Цепное правило . . . . .	196
5.3. Градиенты векторнозначных функций . . . . .	197
5.4. Градиенты матриц . . . . .	204
5.5. Полезные тождества для вычисления градиентов . . . . .	208
5.6. Обратное распространение ошибки и автоматическое дифференцирование . . . . .	209
5.6.1. Градиенты в глубоких нейронных сетях . . . . .	210
5.6.2. Автоматическое дифференцирование . . . . .	211
5.7. Производные высших порядков . . . . .	215
5.8. Линеаризация и ряды Тейлора для нескольких переменных . . . . .	216
5.9. Для дальнейшего чтения . . . . .	222
Упражнения . . . . .	223
<b>Глава 6. Вероятность и распределения . . . . .</b>	<b>225</b>
6.1. Построение вероятностного пространства . . . . .	225
6.1.1. Философские вопросы . . . . .	226
6.1.2. Вероятность и случайные величины . . . . .	228
6.1.3. Статистика . . . . .	231
6.2. Дискретные и непрерывные распределения . . . . .	232
6.2.1. Дискретные вероятности . . . . .	232
6.2.2. Непрерывные вероятности . . . . .	234
6.2.3. Различия дискретных и непрерывных распределений . . . . .	236
6.3. Правило суммы, правило произведения и теорема Байеса . . . . .	238
6.4. Обобщающие статистики и независимость . . . . .	241
6.4.1. Среднее и дисперсия . . . . .	242
6.4.2. Эмпирические среднее и дисперсия . . . . .	247
6.4.3. Три формулы дисперсии . . . . .	248
6.4.4. Суммы и преобразования случайных величин . . . . .	249
6.4.5. Статистическая независимость . . . . .	250
6.4.6. Скалярные произведения случайных величин . . . . .	251
6.5. Гауссово распределение . . . . .	254
6.5.1. Частные и условные распределения — тоже гауссианы . . . . .	255

6.5.2. Произведение гауссовых плотностей . . . . .	258
6.5.3. Суммы и линейные преобразования . . . . .	259
6.5.4. Семплирование из многомерного гауссова распределения . . . . .	262
6.6. Сопряженность и экспоненциальное семейство распределений . . . . .	263
6.6.1. Сопряженность . . . . .	266
6.6.2. Достаточные статистики . . . . .	269
6.6.3. Экспоненциальное семейство распределений . . . . .	270
6.7. Замена переменных / Обратное преобразование . . . . .	274
6.7.1. Метод функций распределения . . . . .	275
6.7.2. Замена переменных . . . . .	277
6.8. Для дальнейшего чтения . . . . .	282
Упражнения . . . . .	283
<b>Глава 7. Непрерывная оптимизация . . . . .</b>	<b>287</b>
7.1. Оптимизация с использованием градиентного спуска . . . . .	290
7.1.1. Размер шага . . . . .	293
7.1.2. Градиентный спуск с импульсом . . . . .	294
7.1.3. Стохастический градиентный спуск . . . . .	295
7.2. Ограниченная оптимизация и множители Лагранжа . . . . .	297
7.3. Выпуклая оптимизация . . . . .	300
7.3.1. Линейное программирование . . . . .	305
7.3.2. Квадратичное программирование . . . . .	307
7.3.3. Преобразование Лежандра — Фенхеля и выпуклое сопряжение . . . . .	308
7.4. Для дальнейшего чтения . . . . .	313
Упражнения . . . . .	314

## ЧАСТЬ II

### Главные задачи машинного обучения

<b>Глава 8. О сочетании модели и данных . . . . .</b>	<b>318</b>
8.1. Данные, модели и обучение . . . . .	318
8.1.1. Данные как векторы . . . . .	319
8.1.2. Модели как функции . . . . .	323
8.1.3. Модели как вероятностные распределения . . . . .	324
8.1.4. Обучение — это нахождение параметров . . . . .	325

8.2. Минимизация эмпирического риска . . . . .	327
8.2.1. Гипотеза класса функций . . . . .	328
8.2.2. Функция потерь для обучения . . . . .	329
8.2.3. Регуляризация для борьбы с переобучением . . . . .	331
8.2.4. Кросс-валидация для оценки производительности обобщения . . . . .	333
8.2.5. Дальнейшее чтение . . . . .	335
8.3. Оценка параметров . . . . .	336
8.3.1. Метод максимального правдоподобия . . . . .	336
8.3.2. Оценка апостериорного максимума . . . . .	339
8.3.3. Обучение модели . . . . .	341
8.3.4. Дополнительное чтение . . . . .	344
8.4. Вероятностные модели и инференс . . . . .	345
8.4.1. Вероятностные модели . . . . .	345
8.4.2. Байесовский инференс . . . . .	346
8.4.3. Модели латентных переменных . . . . .	348
8.4.4. Дальнейшее чтение . . . . .	350
8.5. Направленные графические модели . . . . .	351
8.5.1. Семантика графов . . . . .	352
8.5.2. Условная независимость и d-разбиение . . . . .	355
8.5.3. Дальнейшее чтение . . . . .	357
8.6. Выбор модели . . . . .	358
8.6.1. Вложенная кросс-валидация . . . . .	358
8.6.2. Выбор байесовской модели . . . . .	359
8.6.3. Коэффициент Байеса для сравнения моделей . . . . .	362
8.6.4. Дальнейшее чтение . . . . .	364
<b>Глава 9. Линейная регрессия . . . . .</b>	<b>366</b>
9.1. Постановка задачи . . . . .	368
9.2. Оценка параметров . . . . .	370
9.2.1. Оценка максимального правдоподобия . . . . .	371
9.2.2. Переобучение при линейной регрессии . . . . .	377
9.2.3. Оценка апостериорного максимума . . . . .	380
9.2.4. MAP-оценивание как регуляризация . . . . .	382

9.3. Байесовская линейная регрессия . . . . .	384
9.3.1. Модель . . . . .	384
9.3.2. Априорные предсказания . . . . .	385
9.3.3. Апостериорное распределение . . . . .	387
9.3.4. Апостериорные предсказания . . . . .	389
9.3.5. Вычисление маргинального правдоподобия . . . . .	393
9.4. Максимальное правдоподобие как ортогональная проекция . . . . .	395
9.5. Для дальнейшего чтения . . . . .	397
<b>Глава 10. Снижение размерности с помощью анализа главных компонент . . . . .</b>	<b>400</b>
10.1. Постановка проблемы . . . . .	401
10.2. Перспектива максимальной дисперсии . . . . .	404
10.2.1. Направление с максимальной дисперсией . . . . .	405
10.2.2. M-мерное подпространство с максимальной дисперсией . . . . .	407
10.3. Проекционная перспектива . . . . .	411
10.3.1. Настройка и цели . . . . .	411
10.3.2. Поиск оптимальных координат . . . . .	413
10.3.3. Нахождение базиса главного подпространства . . . . .	415
10.4. Вычисление собственного вектора и приближения низкого ранга . . . . .	419
10.4.1. PCA с использованием матричных приближений низкого ранга . . . . .	420
10.4.2. Практические аспекты . . . . .	421
10.5. PCA в больших размерах . . . . .	422
10.6. Ключевые шаги PCA на практике . . . . .	424
10.7. Латентная переменная . . . . .	427
10.7.1. Генеративный процесс и вероятностная модель . . . . .	428
10.7.2. Правдоподобие и совместное распределение . . . . .	430
10.7.3. Апостериорное распределение . . . . .	431
10.8. Дополнительное чтение . . . . .	432
<b>Глава 11. Оценка плотности с помощью моделей гауссовой смеси . . . . .</b>	<b>438</b>
11.1. Модель гауссовой смеси . . . . .	440
11.2. Изучение параметров с помощью максимального правдоподобия . . . . .	440
11.2.1. Ответственность . . . . .	443
11.2.2. Обновление средних . . . . .	444

11.2.3. Обновление ковариаций	447
11.2.4. Обновление весов смеси	450
11.3. EM-алгоритм	453
11.4. Скрытая перспектива	456
11.4.1. Генеративный процесс и вероятностная модель	456
11.4.2. Правдоподобие	458
11.4.3. Апостериорное распределение	459
11.4.4. Расширение до полного набора данных	459
11.4.5. Еще раз про EM-алгоритм	461
11.5. Дополнительное чтение	461
<b>Глава 12. Классификация методом опорных векторов</b>	<b>464</b>
12.1. Разделяющие гиперплоскости	466
12.2. Прямая задача метода опорных векторов	468
12.2.1. Понятие зазора	469
12.2.2. Нахождение зазора: традиционный способ	471
12.2.3. Почему можно взять зазор, равный 1	473
12.2.4. SVM с мягким зазором: геометрический подход	474
12.2.5. SVM с мягким зазором: подход с использованием функции потерь	476
12.3. Двойственная задача SVM	479
12.3.1. Двойственность и множители Лагранжа	479
12.3.2. Двойственность и выпуклая оболочка	482
12.4. Ядра	484
12.5. Численное решение	487
12.6. Для дальнейшего чтения	489
<b>Библиография</b>	<b>492</b>