

Содержание

От издательства	30
Предисловие	31
Глава 1. Введение	34
1.1. Что такое машинное обучение?.....	34
1.2. Обучение с учителем.....	35
1.2.1. Классификация.....	35
1.2.1.1. Пример: классификация ирисов.....	35
1.2.1.2. Разведочный анализ данных.....	37
1.2.1.3. Обучение классификатора.....	38
1.2.1.4. Минимизация эмпирического риска.....	39
1.2.1.5. Неопределенность.....	41
1.2.1.6. Оценка максимального правдоподобия.....	42
1.2.2. Регрессия.....	43
1.2.2.1. Линейная регрессия.....	44
1.2.2.2. Полиномиальная регрессия.....	45
1.2.2.3. Глубокие нейронные сети.....	46
1.2.3. Переобучение и обобщаемость.....	47
1.2.4. Теорема об отсутствии бесплатных завтраков.....	48
1.3. Обучение без учителя.....	48
1.3.1. Кластеризация.....	49
1.3.2. Обнаружение латентных «факторов изменчивости».....	50
1.3.3. Самостоятельное обучение.....	51
1.3.4. Оценка обучения без учителя.....	52
1.4. Обучение с подкреплением.....	53
1.5. Данные.....	55
1.5.1. Некоторые широко известные наборы изображений.....	55
1.5.1.1. Небольшие наборы изображений.....	55
1.5.1.2. ImageNet.....	56
1.5.2. Некоторые широко известные наборы текстовых данных.....	57
1.5.2.1. Классификация текста.....	58
1.5.2.2. Машинный перевод.....	59
1.5.2.3. Другие задачи типа seq2seq.....	59
1.5.2.4. Языковое моделирование.....	59
1.5.3. Предобработка дискретных входных данных.....	60
1.5.3.1. Унитарное кодирование.....	60
1.5.3.2. Перекрестные произведения признаков.....	60
1.5.4. Предобработка текстовых данных.....	61
1.5.4.1. Модель мешка слов.....	61
1.5.4.2 TF-IDF.....	62

1.5.4.3. Погружения слов.....	63
1.5.4.4. Обработка новых слов	63
1.5.5. Обработка отсутствующих данных	64
1.6. Обсуждение	65
1.6.1. Связь МО с другими дисциплинами	65
1.6.2. Структура книги.....	66
1.6.3. Подводные камни	66

Часть I. ОСНОВАНИЯ.....68

Глава 2. Вероятность: одномерные модели.....69

2.1. Введение.....	69
2.1.1. Что такое вероятность?.....	69
2.1.2. Типы неопределенности.....	70
2.1.3. Вероятность как обобщение логики	70
2.1.3.1. Вероятность события	70
2.1.3.2. Вероятность конъюнкции двух событий	71
2.1.3.3. Вероятность объединения двух событий.....	71
2.1.3.4. Условная вероятность одного события при условии другого	71
2.1.3.5. Независимость событий	72
2.1.3.6. Условная независимость событий	72
2.2. Случайные величины	72
2.2.1. Дискретные случайные величины	72
2.2.2. Непрерывные случайные величины.....	73
2.2.2.1. Функция распределения.....	73
2.2.2.2. Функция плотности распределения	74
2.2.2.3. Квантили.....	75
2.2.3. Множества связанных случайных величин	75
2.2.4. Независимость и условная независимость.....	76
2.2.5. Моменты распределения.....	77
2.2.5.1. Среднее распределения.....	78
2.2.5.2. Дисперсия распределения	78
2.2.5.3. Мода распределения	79
2.2.5.4. Условные моменты.....	80
2.2.6. Ограничения сводных статистик*	81
2.3. Формула Байеса	83
2.3.1. Пример: тестирование на COVID-19.....	84
2.3.2. Пример: парадокс Монти Холла.....	86
2.3.3. Обратные задачи*	88
2.4. Распределение Бернулли и биномиальное распределение.....	89
2.4.1. Определение.....	89
2.4.2. Сигмоидная (логистическая) функция.....	90
2.4.3. Бинарная логистическая регрессия	92
2.5. Категориальное и мультиномиальное распределение	93
2.5.1. Определение.....	93
2.5.2. Функция softmax	94

2.5.3. Многоклассовая логистическая регрессия.....	95
2.5.4. Логарифмирование, суммирование, потенцирование.....	96
2.6. Одномерное гауссово (нормальное) распределение	97
2.6.1. Функция распределения.....	98
2.6.2. Функция плотности вероятности.....	99
2.6.3. Регрессия.....	100
2.6.4. Почему гауссово распределение так широко используется?	101
2.6.5. Дельта-функция Дирака как предельный случай	102
2.7. Другие часто встречающиеся одномерные распределения*	102
2.7.1. Распределение Стьюдента	102
2.7.2. Распределение Коши	104
2.7.3. Распределение Лапласа	105
2.7.4. Бета-распределение.....	105
2.7.5. Гамма-распределение	106
2.7.6. Эмпирическое распределение	107
2.8. Преобразования случайных величин*	108
2.8.1. Дискретный случай	109
2.8.2. Непрерывный случай	109
2.8.3. Обратимые преобразования (биекции)	109
2.8.3.1. Замена переменных: скалярный случай.....	109
2.8.3.2. Замена переменных: многомерный случай.....	110
2.8.4. Моменты линейного преобразования.....	112
2.8.5. Теорема о свертке	113
2.8.6. Центральная предельная теорема.....	115
2.8.7. Аппроксимация Монте-Карло.....	115
2.9. Упражнения.....	116
Глава 3. Вероятность: многомерные модели.....	120
3.1. Совместные распределения нескольких случайных величин.....	120
3.1.1. Ковариация	120
3.1.2. Корреляция	121
3.1.3. Некоррелированные не значит независимые.....	122
3.1.4. Из коррелированности не следует наличие причинно-следственной связи.....	122
3.1.5. Парадокс Симпсона.....	123
3.2. Многомерное гауссово (нормальное) распределение	126
3.2.1. Определение	126
3.2.2. Расстояние Махаланобиса	127
3.2.3. Маргинальные и условные распределения для многомерного нормального распределения*	129
3.2.4. Пример: обусловливание двумерного гауссова распределения.....	130
3.2.5. Пример: подстановка отсутствующих значений*	131
3.3. Линейные гауссовы системы*	132
3.3.1. Формула Байеса для гауссовых распределений	132
3.3.2. Вывод*	133
3.3.3. Пример: вывод неизвестного скаляра.....	134
3.3.4. Пример: вывод неизвестного вектора.....	136

3.3.5. Пример: слияние показаний датчиков.....	137
3.4. Экспоненциальное семейство распределений*	139
3.4.1. Определение.....	139
3.4.2. Пример.....	140
3.4.3. Логарифмическая функция разбиения является производящей функцией кумулянтов.....	141
3.4.4. Вывод максимальной энтропии экспоненциального семейства.....	141
3.5. Смесевые модели.....	142
3.5.1. Модель гауссовой смеси.....	143
3.5.2. Модели бернуллиевой смеси.....	145
3.6. Графовые вероятностные модели*	146
3.6.1. Представление.....	146
3.6.1.1. Пример: оросительная система.....	147
3.6.1.2. Пример: марковская цепь.....	148
3.6.2. Вывод.....	149
3.6.3. Обучение.....	149
3.6.3.1. Блочная нотация.....	150
3.7. Упражнения.....	151
Глава 4. Статистика.....	153
4.1. Введение.....	153
4.2. Оценка максимального правдоподобия (MLE).....	153
4.2.1. Определение.....	154
4.2.2. Обоснование MLE.....	155
4.2.3. Пример: MLE для распределения Бернулли.....	156
4.2.4. Пример: MLE для категориального распределения.....	157
4.2.5. Пример: MLE для одномерного гауссова распределения.....	158
4.2.6. Пример: MLE для многомерного гауссова распределения.....	159
4.2.6.1. MLE среднего.....	159
4.2.6.2. MLE ковариационной матрицы.....	160
4.2.7. Пример: MLE для линейной регрессии.....	161
4.3. Минимизация эмпирического риска (ERM).....	162
4.3.1. Пример: минимизации частоты неправильной классификации.....	163
4.3.2. Суррогатная потеря.....	163
4.4. Другие методы оценивания*	165
4.4.1. Метод моментов.....	165
4.4.1.1. Пример: MOM для одномерного гауссова распределения.....	165
4.4.1.2. Пример: MOM для непрерывного равномерного распределения.....	166
4.4.2. Онлайнное (рекурсивное) оценивание.....	167
4.4.2.1. Пример: рекурсивная MLE среднего гауссова распределения....	167
4.4.2.2. Экспоненциально взвешенное скользящее среднее.....	167
4.5. Регуляризация.....	169
4.5.1. Пример: оценка MAP для распределения Бернулли.....	170
4.5.2. Пример: оценка MAP для многомерного гауссова распределения* ...	171
4.5.2.1. Оценка усадки.....	171
4.5.3. Пример: уменьшение весов.....	172

4.5.4. Подбор регуляризатора с помощью контрольного набора	173
4.5.5. Перекрестная проверка	174
4.5.5.1. Правило одной стандартной ошибки	175
4.5.5.2. Пример: гребневая регрессия	176
4.5.6. Ранняя остановка	176
4.5.7. Больше данных	177
4.6. Байесовские статистики*	178
4.6.1. Сопряженные априорные распределения	179
4.6.2. Бета-биномиальная модель	180
4.6.2.1. Правдоподобие Бернулли	180
4.6.2.2. Биномиальное правдоподобие	180
4.6.2.3. Априорное распределение	181
4.6.2.4. Апостериорное распределение	181
4.6.2.5. Пример	181
4.6.2.6. Апостериорная мода (оценка MAP)	182
4.6.2.7. Апостериорное среднее	183
4.6.2.8. Апостериорная дисперсия	183
4.6.2.9. Апостериорное прогнозное распределение	184
4.6.2.10. Маргинальное правдоподобие	187
4.6.2.11. Смеси сопряженных априорных распределений	187
4.6.3. Дирихле-мультиномиальная модель	189
4.6.3.1. Правдоподобие	189
4.6.3.2. Априорное распределение	189
4.6.3.3. Апостериорное распределение	191
4.6.3.4. Апостериорное прогнозное распределение	192
4.6.3.5. Маргинальное правдоподобие	192
4.6.4. Гауссова-гауссова модель	193
4.6.4.1. Одномерный случай	193
4.6.4.2. Многомерный случай	195
4.6.5. За пределами сопряженных априорных распределений	196
4.6.5.1. Неинформативные априорные распределения	197
4.6.5.2. Иерархические априорные распределения	197
4.6.5.3. Эмпирические априорные распределения	197
4.6.6. Байесовские доверительные интервалы	198
4.6.7. Байесовское машинное обучение	200
4.6.7.1. Подстановочная аппроксимация	201
4.6.7.2. Пример: скалярный вход, бинарный выход	201
4.6.7.3. Пример: бинарный вход, скалярный выход	203
4.6.7.4. Вертикальное масштабирование	205
4.6.8. Вычислительные трудности	205
4.6.8.1. Сеточная аппроксимация	206
4.6.8.2. Квадратичная аппроксимация (Лапласа)	206
4.6.8.3. Вариационная аппроксимация	207
4.6.8.4. Аппроксимация методом Монте-Карло по схеме марковских цепей	208
4.7. Частотная статистика*	208
4.7.1. Выборочное распределение	209

4.7.2. Гауссова аппроксимация выборочного распределения MLE.....	210
4.7.3. Бутстрэпная аппроксимация выборочного распределения любого оценителя.....	211
4.7.3.1. Бутстрэп – апостериорное распределение «для бедных».....	211
4.7.4. Доверительные интервалы.....	212
4.7.5. Предостережения: доверительные интервалы и байесовские доверительные интервалы не одно и то же.....	214
4.7.6. Компромисс между смещением и дисперсией.....	215
4.7.6.1. Смещение оценки.....	215
4.7.6.2. Дисперсия оценки.....	216
4.7.6.3. Компромисс между смещением и дисперсией.....	216
4.7.6.4. Пример: оценка MAP среднего гауссова распределения.....	217
4.7.6.5. Пример: оценка MAP для линейной регрессии.....	218
4.7.6.6. Применение компромисса между смещением и дисперсией для классификации.....	220
4.8. Упражнения.....	220
Глава 5. Теория принятия решений.....	225
5.1. Байесовская теория принятия решений.....	225
5.1.1. Основы.....	225
5.1.2. Проблемы классификации.....	227
5.1.2.1. Бинарная потеря.....	228
5.1.2.2. Классификация с учетом стоимости.....	228
5.1.2.3. Классификация с возможностью отклонения примера.....	229
5.1.3. ROC-кривые.....	230
5.1.3.1. Матрицы неточностей классификации.....	230
5.1.3.2. Обобщение ROC-кривой в виде скаляра.....	233
5.1.3.3. Несбалансированность классов.....	233
5.1.4. Кривые точность–полнота.....	233
5.1.4.1. Вычисление точности и полноты.....	234
5.1.4.2. Обобщение кривых точность–полнота в виде скаляра.....	234
5.1.4.3. F-мера.....	235
5.1.4.4. Несбалансированность классов.....	235
5.1.5. Задачи регрессии.....	236
5.1.5.1. ℓ_2 -потеря.....	236
5.1.5.2. ℓ_1 -потеря.....	237
5.1.5.3. Функция потерь Хьюбера.....	237
5.1.6. Задачи вероятностного предсказания.....	238
5.1.6.1. Расхождение КЛ, перекрестная энтропия и логарифмическая потеря.....	238
5.1.6.2. Правила верной оценки.....	239
5.2. Байесовская проверка гипотез.....	240
5.2.1. Пример: проверка симметричности монеты.....	241
5.2.2. Байесовский выбор модели.....	242
5.2.2.1. Пример: полиномиальная регрессия.....	243
5.2.3. Бритва Оккама.....	244
5.2.4. Связь между перекрестной проверкой и маргинальным правдоподобием.....	246

5.2.5. Информационные критерии.....	246
5.2.5.1. Байесовский информационный критерий (BIC)	247
5.2.5.2. Информационный критерий Акаике	247
5.2.5.3. Минимальная длина описания (MDL).....	248
5.3. Частотная теория принятий решений	248
5.3.1. Вычисление риска оценки.....	248
5.3.1.1. Пример	249
5.3.1.2. Байесовский риск	250
5.3.1.3. Максимальный риск	251
5.3.2. Состоятельные оценки.....	251
5.3.3. Допустимые оценки	252
5.4. Минимизация эмпирического риска.....	253
5.4.1. Эмпирический риск.....	253
5.4.1.1. Ошибка аппроксимации и ошибка оценивания	254
5.4.1.2. Регуляризованный риск.....	255
5.4.2. Структурный риск.....	255
5.4.3. Перекрестная проверка	256
5.4.4. Статистическая теория обучения*	257
5.4.4.1. Нахождение границы ошибки обобщения	257
5.4.4.2. VC-размерность	258
5.5. Частотная проверка гипотез*	258
5.5.1. Критерий отношения правдоподобия.....	259
5.5.1.1. Пример: сравнение гауссовых средних	259
5.5.1.2. Простые и сложные гипотезы.....	260
5.5.2. Проверка значимости нулевой гипотезы	260
5.5.3. p-значения	261
5.5.4. О вреде p-значений.....	261
5.5.5. Почему же не все исповедуют байесовский подход?	264
5.6. Упражнения.....	266
Глава 6. Теория информации.....	268
6.1. Энтропия	268
6.1.1. Энтропия дискретных случайных величин	268
6.1.2. Перекрестная энтропия	271
6.1.3. Совместная энтропия.....	271
6.1.4. Условная энтропия.....	272
6.1.5. Перплексия	273
6.1.6. Дифференциальная энтропия непрерывных случайных величин*	274
6.1.6.1. Пример: энтропия гауссова распределения.....	274
6.1.6.2. Связь с дисперсией.....	275
6.1.6.3. Дискретизация.....	275
6.2. Относительная энтропия (расхождение KL)*	275
6.2.1. Определение.....	276
6.2.2. Интерпретация.....	276
6.2.3. Пример: расхождение КЛ между двумя гауссовыми распределениями.....	276

6.2.4. Неотрицательность расхождения КЛ.....	277
6.2.5. Расхождение КЛ и оценка максимального правдоподобия.....	278
6.2.6. Прямое и обратное расхождение КЛ.....	279
6.3. Взаимная информация*.....	280
6.3.1. Определение.....	280
6.3.2. Интерпретация.....	280
6.3.3. Пример.....	282
6.3.4. Условная взаимная информация.....	282
6.3.5. Взаимная информация как «обобщенный коэффициент корреляции».....	283
6.3.6. Нормированная взаимная информация.....	284
6.3.7. Максимальный коэффициент информации.....	285
6.3.8. Неравенство обработки данных.....	287
6.3.9. Достаточные статистики.....	288
6.3.10. Неравенство Фано*.....	288
6.4. Упражнения.....	289
Глава 7. Линейная алгебра.....	292
7.1. Введение.....	292
7.1.1. Обозначения.....	292
7.1.1.1. Векторы.....	292
7.1.1.2. Матрицы.....	293
7.1.1.3. Тензоры.....	294
7.1.2. Векторные пространства.....	295
7.1.2.1. Сложение векторов и умножение вектора на скаляр.....	295
7.1.2.2. Линейная независимость, линейная оболочка и базисы.....	296
7.1.2.3. Линейные отображения и матрицы.....	296
7.1.2.4. Образ и ядро матрицы.....	297
7.1.2.5. Линейная проекция.....	297
7.1.3. Нормы вектора и матрицы.....	298
7.1.3.1. Нормы вектора.....	298
7.1.3.2. Нормы матрицы.....	299
7.1.4. Свойства матриц.....	300
7.1.4.1. След квадратной матрицы.....	300
7.1.4.2. Определитель квадратной матрицы.....	300
7.1.4.3. Ранг матрицы.....	301
7.1.4.4. Числа обусловленности.....	301
7.1.5. Специальные типы матриц.....	303
7.1.5.1. Диагональная матрица.....	303
7.1.5.2. Треугольные матрицы.....	304
7.1.5.3. Положительно определенные матрицы.....	304
7.1.5.4. Ортогональные матрицы.....	305
7.2. Умножение матриц.....	306
7.2.1. Умножение векторов.....	307
7.2.2. Произведение матрицы на вектор.....	307
7.2.3. Произведение матриц.....	308
7.2.4. Приложение: манипулирование матрицами данных.....	310

7.2.4.1. Суммирование срезов матрицы	310
7.2.4.2. Масштабирование строк и столбцов матрицы.....	311
7.2.4.3. Матрица сумм квадратов и матрица рассеяния	311
7.2.4.4. Матрица Грама	312
7.2.4.5. Матрица расстояний	313
7.2.5. Произведения Кронекера*	313
7.2.6. Суммирование Эйнштейна*	314
7.3. Обращение матриц	315
7.3.1. Обращение квадратной матрицы.....	315
7.3.2. Дополнения Шура*	316
7.3.3. Лемма об обращении матрицы*	317
7.3.4. Лемма об определителе матрицы*	318
7.3.5. Приложение: вывод условных распределений для многомерного гауссова распределения	319
7.4. Спектральное разложение	320
7.4.1. Основные сведения.....	320
7.4.2. Диагонализация	321
7.4.3. Собственные значения и собственные векторы симметричных матриц	322
7.4.3.1. Проверка на положительную определенность.....	322
7.4.4. Геометрия квадратичных форм.....	323
7.4.5. Стандартизация и отбеливание данных.....	323
7.4.6. Степенной метод.....	324
7.4.7. Понижение порядка	326
7.4.8. Собственные векторы оптимизируют квадратичные формы.....	326
7.5. Сингулярное разложение (SVD)	327
7.5.1. Основные сведения.....	327
7.5.2. Связь между сингулярным и спектральным разложением.....	328
7.5.3. Псевдообратная матрица.....	329
7.5.4. SVD для образа и ядра матрицы*	330
7.5.5. Усеченное сингулярное разложение	331
7.6. Другие матричные разложения*	332
7.6.1. LU-разложение	332
7.6.2. QR-разложение	333
7.6.3. Разложение Холецки	334
7.6.3.1. Приложение: выборка из многомерного гауссова распределения	334
7.7. Решение систем линейных уравнений*	335
7.7.1. Решение квадратных систем	336
7.7.2. Решение недоопределенных систем (оценка по наименьшей норме).....	336
7.7.3. Решение переопределенных систем (оценка по методу наименьших квадратов)	338
7.8. Матричное исчисление.....	339
7.8.1. Производные	339
7.8.2. Градиенты	340
7.8.3. Производная по направлению	340

7.8.4. Полная производная*	341
7.8.5. Якобиан	341
7.8.5.1. Умножение якобиана на вектор	342
7.8.5.2. Якобиан композиции	342
7.8.6. Гессиан	342
7.8.7. Градиенты часто встречающихся функций	343
7.8.7.1. Функции, отображающие скаляры в скаляры	343
7.8.7.2. Функции, отображающие векторы в скаляры	343
7.8.7.3. Функции, отображающие матрицы в скаляры	344
7.9. Упражнения	345

Глава 8. Оптимизация 346

8.1. Введение	346
8.1.1. Локальная и глобальная оптимизация	346
8.1.1.1. Условия оптимальности для локальных и глобальных оптимумов	347
8.1.2. Условная и безусловная оптимизация	348
8.1.3. Выпуклая и невыпуклая оптимизация	349
8.1.3.1. Выпуклые множества	349
8.1.3.2. Выпуклые функции	350
8.1.3.3. Характеристика выпуклых функций	351
8.1.3.4. Сильно выпуклые функции	352
8.1.4. Гладкая и негладкая оптимизация	353
8.1.4.1. Субградиенты	354
8.2. Методы первого порядка	355
8.2.1. Направление спуска	356
8.2.2. Размер шага (скорость обучения)	356
8.2.2.1. Постоянный размер шага	356
8.2.2.2. Линейный поиск	358
8.2.3. Скорость сходимости	359
8.2.4. Метод импульса	360
8.2.4.1. Импульс	360
8.2.4.2. Импульс Нестерова	361
8.3. Методы второго порядка	362
8.3.1. Метод Ньютона	362
8.3.2. BFGS и другие квазиньютоновские методы	364
8.3.3. Методы на основе доверительных областей	365
8.4. Стохастический градиентный спуск	366
8.4.1. Приложение к задачам с конечной суммой	367
8.4.2. Пример: СГС для обучения модели линейной регрессии	368
8.4.3. Выбор размера шага (скорости обучения)	369
8.4.4. Итеративное усреднение	371
8.4.5. Уменьшение дисперсии*	372
8.4.5.1. SVRG	372
8.4.5.2. SAGA	373
8.4.5.3. Применение в глубоком обучении	373
8.4.6. Предобусловленный СГС	374

8.4.6.1. AdaGrad	374
8.4.6.2. RMSProp и AdaDelta.....	375
8.4.6.3. Adam	376
8.4.6.4. Проблемы, связанные с адаптивной скоростью обучения	376
8.4.6.5. Недиagonальные матрицы преобусловливания	377
8.5. Условная оптимизация.....	377
8.5.1. Множители Лагранжа.....	378
8.5.1.1. Пример: двумерная квадратичная целевая функция с одним линейным ограничением в виде равенства.....	379
8.5.2. Условия Каруша–Куна–Таккера	380
8.5.3. Линейное программирование	381
8.5.3.1. Симплекс-метод	382
8.5.3.2. Приложения.....	382
8.5.4. Квадратичное программирование.....	382
8.5.4.1. Пример: квадратичная целевая функция в двумерном случае с линейными ограничениями в виде равенств	383
8.5.4.2. Приложения.....	384
8.5.5. Смешанно-целочисленное линейное программирование*	384
8.6. Проксимальный градиентный метод*	384
8.6.1. Спроецированный градиентный спуск.....	385
8.6.2. Проксимальный оператор для регуляризатора по норме ℓ_1	387
8.6.3. Применение проксимального оператора в случае квантования	388
8.6.4. Инкрементные (онлайнные) проксимальные методы	389
8.7. Граничная оптимизация*	389
8.7.1. Общий алгоритм	389
8.7.2. EM-алгоритм	391
8.7.2.1. Нижняя граница.....	392
8.7.2.2. E-шаг	392
8.7.2.3. M-шаг	393
8.7.3. Пример: EM-алгоритм для смеси гауссовых распределений.....	394
8.7.3.1. E-шаг	394
8.7.3.2. M-шаг	394
8.7.3.3. Пример	395
8.7.3.4. Оценка MAP	395
8.7.3.5. Невыпуклость NLL	398
8.8. Оптимизация черного ящика и оптимизация без использования производных.....	399
8.9. Упражнения.....	399

Часть II. ЛИНЕЙНЫЕ МОДЕЛИ..... 400

Глава 9. Линейный дискриминантный анализ..... 401

9.1. Введение	401
9.2. Гауссов дискриминантный анализ	401
9.2.1. Квадратичные решающие границы.....	402
9.2.2. Линейные решающие границы	403

9.2.3. Связь между ЛДА и логистической регрессией.....	403
9.2.4. Обучение модели	405
9.2.4.1. Связанные ковариационные матрицы	406
9.2.4.2. Диагональные ковариационные матрицы	406
9.2.4.3. Оценка MAP.....	406
9.2.5. Классификатор по ближайшему центру.....	407
9.2.6. Линейный дискриминантный анализ Фишера*	407
9.2.6.1. Нахождение оптимального одномерного направления	409
9.2.6.2. Обобщение на большую размерность и несколько классов	411
9.3. Наивные байесовские классификаторы	412
9.3.1. Примеры моделей.....	413
9.3.2. Обучение модели	413
9.3.3. Байесовская интерпретация наивной байесовской модели	415
9.3.4. Связь между наивной байесовской моделью и логистической регрессией.....	416
9.4. Порождающие и дискриминантные классификаторы.....	417
9.4.1. Преимущества дискриминантных классификаторов	417
9.4.2. Преимущества порождающих классификаторов.....	418
9.4.3. Обработка отсутствующих признаков.....	419
9.5. Упражнения.....	419
Глава 10. Логистическая регрессия	420
10.1. Введение	420
10.2. Бинарная логистическая регрессия.....	420
10.2.1. Линейные классификаторы	421
10.2.2. Нелинейные классификаторы	422
10.2.3. Оценка максимального правдоподобия	423
10.2.3.1. Целевая функция	423
10.2.3.2. Оптимизация целевой функции	424
10.2.3.3. Вывод градиента.....	425
10.2.3.4. Вывод гессиана	426
10.2.4. Стохастический градиентный спуск.....	427
10.2.5. Алгоритм перцептрона	427
10.2.6. Метод наименьших квадратов с итеративным пересчетом весов.....	428
10.2.7. Оценка MAP	430
10.2.8. Стандартизация	431
10.3. Мультиномиальная логистическая регрессия	432
10.3.1. Линейные и нелинейные классификаторы	433
10.3.2. Оценка максимального правдоподобия	433
10.3.2.1. Целевая функция	434
10.3.2.2. Оптимизация целевой функции	434
10.3.2.3. Вывод градиента.....	434
10.3.2.4. Вывод гессиана	435
10.3.3. Градиентная оптимизация	436
10.3.4. Граничная оптимизация.....	436
10.3.5. Оценка MAP	438

10.3.6. Классификаторы максимальной энтропии	439
10.3.7. Иерархическая классификация.....	440
10.3.8. Работа с большим числом классов	440
10.3.8.1. Иерархическая softmax-модель.....	441
10.3.8.2. Несбалансированность классов и длинный хвост.....	441
10.4. Робастная логистическая регрессия*	443
10.4.1. Смесевая модель правдоподобия.....	443
10.4.2. Дважды смягченная потеря	444
10.5. Байесовская логистическая регрессия*	447
10.5.1. Аппроксимация Лапласа	447
10.5.2. Аппроксимация апостериорного прогнозного распределения	449
10.5.2.1. Аппроксимация Монте-Карло.....	451
10.5.2.2. Пробит-аппроксимация	451
10.6. Упражнения.....	452
Глава 11. Линейная регрессия	455
11.1. Введение	455
11.2. Линейная регрессия по методу наименьших квадратов	455
11.2.1. Терминология.....	455
11.2.2. Оценивание по методу наименьших квадратов.....	457
11.2.2.1. Обыкновенный метод наименьших квадратов	457
11.2.2.2. Геометрическая интерпретация метода наименьших квадратов	458
11.2.2.3. Алгоритмические проблемы	460
11.2.2.4. Метод взвешенных наименьших квадратов	461
11.2.3. Другие подходы к вычислению MLE	461
11.2.3.1. Нахождение смещения и углового коэффициента по отдельности.....	461
11.2.3.2. Простая линейная регрессия (одномерные входные данные)	462
11.2.3.3. Частная регрессия	462
11.2.3.4. Рекурсивное вычисление MLE.....	462
11.2.3.5. Вывод MLE с порождающей точки зрения	464
11.2.3.6. Вывод MLE для σ^2	465
11.2.4. Измерение степени согласия оценки	465
11.2.4.1. Графики невязок.....	465
11.2.4.2. Точность предсказания и R^2	466
11.3. Гребневая регрессия	467
11.3.1. Вычисление оценки MAP.....	467
11.3.1.1. Решение с использованием QR-разложения.....	468
11.3.1.2. Решение с использованием сингулярного разложения	469
11.3.2. Связь между гребневой регрессией и PCA.....	469
11.3.3. Выбор силы регуляризатора	471
11.4. Регрессия lasso	471
11.4.1. Оценка MAP с априорным распределением Лапласа (ℓ_1 -регуляризация).....	472
11.4.2. Почему ℓ_1 -регуляризация дает разреженные решения?	473

11.4.3. Жесткие и мягкие пороги	474
11.4.4. Путь регуляризации	476
11.4.5. Сравнение методов наименьших квадратов, lasso, гребневой регрессии и выбора подмножеств	478
11.4.6. Согласованность выбора переменных	479
11.4.7. Групповое lasso	481
11.4.7.1. Приложения	481
11.4.7.2. Штрафование по норме ℓ_2	482
11.4.7.3. Штрафование по норме ℓ_∞	482
11.4.7.4. Пример	483
11.4.8. Эластичная сеть (комбинация гребневой регрессии и lasso)	484
11.4.9. Алгоритмы оптимизации	485
11.4.9.1. Покоординатный спуск	485
11.4.9.2. Спроецированный градиентный спуск	486
11.4.9.3. Проксимальный градиентный спуск	486
11.4.9.4. LARS	486
11.5. Регрессионные сплайны*	487
11.5.1. В-сплайны в качестве базисных функций	488
11.5.2. Обучение линейно модели с помощью сплайнового базиса	489
11.5.3. Сглаживающие сплайны	490
11.5.4. Обобщенные аддитивные модели	490
11.6. Робастная линейная регрессия*	491
11.6.1. Правдоподобие Лапласа	491
11.6.1.1. Вычисление MLE методами линейного программирования	492
11.6.2. t-правдоподобие Стьюдента	493
11.6.3. Функция потерь Хьюбера	493
11.6.4. RANSAC	494
11.7. Байесовская линейная регрессия*	494
11.7.1. Априорные распределения	494
11.7.2. Апостериорные распределения	495
11.7.3. Пример	495
11.7.4. Вычисление апостериорного прогнозного распределения	497
11.7.5. Преимущество центрирования	498
11.7.6. Мультиколлинеарность	499
11.7.7. Автоматическое определение релевантности (ARD)*	501
11.8. Упражнения	502
Глава 12. Обобщенные линейные модели*	505
12.1. Введение	505
12.2. Примеры	506
12.2.1. Линейная регрессия	506
12.2.2. Биномиальная регрессия	506
12.2.3. Регрессия Пуассона	507
12.3. GLM с неканоническими функциями связи	508
12.4. Оценка максимального правдоподобия	509
12.5. Рабочий пример: предсказание обращений за страховыми выплатами	510

Часть III. ГЛУБОКИЕ НЕЙРОННЫЕ СЕТИ	513
Глава 13. Нейронные сети для структурированных данных	514
13.1. Введение	514
13.2. Многослойные перцептроны (МСП)	516
13.2.1. Задача XOR	516
13.2.2. Дифференцируемые МСП	517
13.2.3. Функции активации	518
13.2.4. Примеры моделей	519
13.2.4.1. МСП для классификации двумерных данных по двум категориям	519
13.2.4.2. МСП для классификации изображений	520
13.2.4.3. МСП для классификации текстов	522
13.2.4.4. МСП для гетероскедастической регрессии	523
13.2.5. Важность глубины	524
13.2.6. Революция глубокого обучения	525
13.2.7. Связи с биологией	526
13.3. Обратное распространение	529
13.3.1. Прямой и обратный режим дифференцирования	530
13.3.2. Дифференцирование в обратном режиме для многослойных перцептронов	531
13.3.3. Произведение вектора на якобиан для типичных слоев	533
13.3.3.1. Слой перекрестной энтропии	533
13.3.3.2. Поэлементная нелинейность	534
13.3.3.3. Линейный слой	535
13.3.3.4. Соберем все вместе	536
13.3.4. Графы вычислений	536
13.4. Обучение нейронных сетей	538
13.4.1. Настройка скорости обучения	539
13.4.2. Исчезающие и взрывные градиенты	539
13.4.3. Функции активации без насыщения	540
13.4.3.1. ReLU	542
13.4.3.2. ReLU без насыщения	542
13.4.3.3. Другие варианты	543
13.4.4. Остаточные связи	544
13.4.5. Инициализация параметров	545
13.4.5.1. Эвристические схемы инициализации	545
13.4.5.2. Инициализации, управляемые данными	546
13.4.6. Параллельное обучение	546
13.5. Регуляризация	548
13.5.1. Ранняя остановка	548
13.5.2. Уменьшение весов	548
13.5.3. Разреженные ГНС	548
13.5.4. Прореживание	549
13.5.5. Байесовские нейронные сети	551
13.5.6. Эффекты регуляризации, порождаемые стохастическим градиентным спуском*	551

13.6. Другие виды сетей прямого распространения*	553
13.6.1. Сети радиально-базисных функций	553
13.6.1.1. RBF-сеть для регрессии	554
13.6.1.2. RBF-сеть для классификации	554
13.6.2. Смесь экспертов	555
13.6.2.1. Смесь линейных экспертов	558
13.6.2.2. Сети на основе смеси моделей разной плотности	558
13.6.2.3. Иерархические смеси экспертов	559
13.7. Упражнения	559

Глава 14. Нейронные сети для изображений 561

14.1. Введение	561
14.2. Наиболее употребительные слои	563
14.2.1. Сверточные слои	563
14.2.1.1. Свертка в одномерном случае	563
14.2.1.2. Свертка в двумерном случае	564
14.2.1.3. Свертка как умножение матрицы на вектор	565
14.2.1.4. Граничные условия и дополнение	566
14.2.1.5. Свертка с шагом	568
14.2.1.6. Несколько входных и выходных каналов	568
14.2.1.7. Свертка 1×1 (поточечная)	569
14.2.2. Пулинговые слои	569
14.2.3. Соберем все вместе	571
14.2.4. Слои нормировки	571
14.2.4.1. Пакетная нормировка	572
14.2.4.2. Другие виды слоя нормировки	573
14.2.4.3. Сети без нормировки	575
14.3. Распространенные архитектуры классификации изображений	575
14.3.1. LeNet	575
14.3.2. AlexNet	577
14.3.3. GoogLeNet	578
14.3.4. ResNet	579
14.3.5. DenseNet	581
14.3.6. Поиск архитектуры нейронной сети	581
14.4. Другие формы свертки*	582
14.4.1. Дырявая свертка	582
14.4.2. Транспонированная свертка	583
14.4.3. Пространственная раздельная свертка	584
14.5. Решение других дискриминантных задач компьютерного зрения с помощью СНС*	585
14.5.1. Аннотирование изображений	586
14.5.2. Определение объектов	586
14.5.3. Сегментация экземпляров	588
14.5.4. Семантическая сегментация	589
14.5.5. Оценивание позы человека	590
14.6. Генерирование изображений посредством инвертирования СНС*	591

14.6.1. Преобразование обученного классификатора в порождающую модель.....	592
14.6.2. Априорные распределения изображений.....	592
14.6.2.1. Гауссово априорное распределения	593
14.6.2.2. Априорное распределение на основе полной вариации	594
14.6.3. Визуализация признаков, обученных с помощью СНС	595
14.6.4. Deep Dream.....	595
14.6.5. Нейронный перенос стиля	597
14.6.5.1. Как это работает	598
14.6.5.2. Ускорение метода	600
Глава 15. Нейронные сети для последовательностей.....	602
15.1. Введение	602
15.2. Рекуррентные нейронные сети (РНС).....	602
15.2.1. Vec2Seq (генерирование последовательностей)	602
15.2.1.1. Модели	603
15.2.1.2. Приложения.....	604
15.2.2. Seq2Vec (классификация последовательностей).....	606
15.2.3. Seq2Seq (трансляция последовательностей)	607
15.2.3.1. Выровненный случай.....	607
15.2.3.2. Невыровненный случай	608
15.2.4. Принуждение со стороны учителя	609
15.2.5. Обратное распространение во времени	610
15.2.6. Исчезающие и взрывные градиенты	612
15.2.7. Вентильная и долгосрочная память	612
15.2.7.1. Управляемые рекуррентные блоки (GRU).....	612
15.2.7.2. Долгая краткосрочная память (LSTM)	613
15.2.8. Лучевой поиск	615
15.3. Одномерные СНС	617
15.3.1. Применение одномерных СНС для классификации последовательностей	618
15.3.2. Применение каузальных одномерных СНС для генерирования последовательностей	618
15.4. Модель внимания.....	620
15.4.1. Механизм внимания как мягкий поиск в словаре	620
15.4.2. Ядерная регрессия как непараметрическое внимание	622
15.4.3. Параметрическое внимание	623
15.4.4. Модель Seq2Seq с вниманием	624
15.4.5. Модель Seq2Vec с вниманием (классификация текста).....	626
15.4.6. Модель Seq+Seq2Vec с вниманием (классификация пар предложений).....	627
15.4.7. Жесткое внимание по сравнению с мягким	629
15.5. Трансформеры.....	629
15.5.1. Самовнимание	630
15.5.2. Многоголовое внимание	632
15.5.3. Позиционное кодирование	632
15.5.4. Соберем все вместе.....	634

15.5.5. Сравнение трансформеров, СНС и РНС.....	636
15.5.6. Применение трансформеров для изображений*.....	636
15.5.7. Другие варианты трансформеров*.....	638
15.6. Эффективные трансформеры*.....	639
15.6.1. Фиксированные необучаемые локализованные паттерны внимания.....	639
15.6.2. Обучаемые паттерны разреженного внимания.....	640
15.6.3. Методы с добавлением памяти и рекуррентные методы.....	640
15.6.4. Низкоранговые и ядерные методы.....	640
15.7. Языковые модели и обучение представлений без учителя.....	643
15.7.1. ELMo.....	643
15.7.2. BERT.....	644
15.7.2.1. Замаскированная языковая модель.....	645
15.7.2.2. Задача предсказания следующего предложения.....	645
15.7.2.3. Дообучение BERT для приложений NLP.....	647
15.7.3. GPT.....	649
15.7.3.1. Приложения GPT.....	649
15.7.4. T5.....	649
15.7.5. Обсуждение.....	650

Часть IV. НЕПАРАМЕТРИЧЕСКИЕ МОДЕЛИ..... 652

Глава 16. Методы на основе эталонов..... 653

16.1. Классификация методом К ближайших соседей (KNN).....	653
16.1.1. Пример.....	654
16.1.2. Проклятие размерности.....	655
16.1.3. Снижение требований к скорости и памяти.....	656
16.1.4. Распознавание открытого множества.....	657
16.1.4.1. Онлайнное обучение, обнаружение посторонних и распознавание открытого множества.....	657
16.1.4.2. Другие задачи открытого мира.....	658
16.2. Обучение метрик.....	658
16.2.1. Линейные и выпуклые методы.....	659
16.2.1.1. Метод ближайших соседей с большим зазором.....	659
16.2.1.2. Анализ компонентов соседства.....	660
16.2.1.3. Анализ латентных совпадений.....	660
16.2.2. Глубокое обучение метрики.....	661
16.2.3. Потери классификации.....	662
16.2.4. Потери ранжирования.....	662
16.2.4.1. Попарная (сопоставительная) потеря и сиамские сети.....	663
16.2.4.2. Триpletная потеря.....	663
16.2.4.3. N-парная потеря.....	664
16.2.5. Ускорение оптимизации потери ранжирования.....	665
16.2.5.1. Методы на основе расширения.....	665
16.2.5.2. Методы на основе представителей.....	665
16.2.5.3. Оптимизация верхней границы.....	666

16.2.6. Другие приемы глубокого обучения метрики.....	668
16.3. Ядерные оценки плотности.....	669
16.3.1. Ядра плотности	669
16.3.2. Оконная оценка плотности Парзена	670
16.3.3. Как выбирать полосу пропускания	672
16.3.4. От KDE к KNN-классификации	672
16.3.5. Ядерная регрессия	673
16.3.5.1. Оценка среднего Надарая–Ватсона.....	673
16.3.5.2. Оценка дисперсии.....	675
16.3.5.3. Локально взвешенная регрессия.....	675
Глава 17. Ядерные методы*	676
17.1. Ядра Мерсера.....	676
17.1.1. Теорема Мерсера	678
17.1.2. Некоторые популярные ядра Мерсера.....	678
17.1.2.1. Стационарные ядра для вещественных векторов	678
17.1.2.2. Создание новых ядер из существующих.....	681
17.1.2.3. Комбинирование ядер с помощью сложения и умножения	682
17.1.2.4. Ядра для структурированных входов	683
17.2. Гауссовы процессы	683
17.2.1. Незашумленные наблюдения.....	684
17.2.2. Зашумленные наблюдения.....	685
17.2.3. Сравнение с ядерной регрессией	686
17.2.4. Пространство весов и пространство функций.....	687
17.2.5. Численные проблемы	688
17.2.6. Оценивание параметров ядра.....	688
17.2.6.1. Эмпирическая байесовская оценка	689
17.2.6.2. Байесовский вывод.....	691
17.2.7. Применение гауссовых процессов для классификации	692
17.2.8. Связи с глубоким обучением.....	694
17.2.9. Масштабирование ГП на большие наборы данных	694
17.2.9.1. Разреженные аппроксимации	694
17.2.9.2. Распараллеливание с использованием структуры ядерной матрицы.....	694
17.2.9.3. Аппроксимация случайными признаками.....	695
17.3. Метод опорных векторов	696
17.3.1. Классификаторы с широким зазором.....	697
17.3.2. Двойственная задача	699
17.3.3. Классификаторы с мягким зазором	701
17.3.4. Ядерный трюк.....	702
17.3.5. Преобразование выходов SVM в вероятности.....	703
17.3.6. Связь с логистической регрессией	704
17.3.7. Многоклассовая классификация с применением SVM.....	705
17.3.8. Как выбирать регуляризатор C	706
17.3.9. Ядерная гребневая регрессия.....	707
17.3.10. Применение SVM для регрессии.....	708
17.4. Метод разреженных векторов	711

17.4.1. Метод релевантных векторов.....	711
17.4.2. Сравнение разреженных и плотных ядерных методов	711
17.5. Упражнения	715

Глава 18. Деревья, леса, бэггинг и бустинг 716

18.1. Деревья классификации и регрессии.....	716
18.1.1. Определение модели.....	716
18.1.2. Обучение модели	717
18.1.3. Регуляризация	719
18.1.4. Обработка отсутствующих входных признаков	720
18.1.5. Плюсы и минусы	720
18.2. Ансамблевое обучение	721
18.2.1. Стековое обобщение	722
18.2.2. Ансамблевое обучение не то же, что байесовское усреднение моделей	722
18.3. Бэггинг	723
18.4. Случайные леса.....	724
18.5. Бустинг.....	725
18.5.1. Прямое поэтапное аддитивное моделирование.....	726
18.5.2. Квадратичная потеря и бустинг наименьших квадратов.....	727
18.5.3. Экспоненциальная потеря и AdaBoost	727
18.5.4. LogitBoost	731
18.5.5. Градиентный бустинг	732
18.5.5.1. Градиентный бустинг деревьев.....	734
18.5.5.2. XGBoost	734
18.6. Интерпретация ансамблей деревьев	736
18.6.1. Важность признаков.....	736
18.6.2. Графики частичной зависимости.....	738

Часть V. ЗА ПРЕДЕЛАМИ ОБУЧЕНИЯ С УЧИТЕЛЕМ 739

Глава 19. Обучение при меньшем числе помеченных примеров 740

19.1. Приращение данных.....	740
19.1.1. Примеры.....	740
19.1.2. Теоретическое обоснование.....	741
19.2. Перенос обучения	742
19.2.1. Дообучение	742
19.2.2. Адаптеры.....	744
19.2.3. Предобучение с учителем.....	745
19.2.4. Предобучение без учителя (самостоятельное обучение)	746
19.2.4.1. Задачи подстановки.....	747
19.2.4.2. Замещающие задачи.....	748
19.2.4.3. Сопоставительные задачи.....	748
19.2.4.4. SimCLR.....	748
19.2.4.5. CLIP	751

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru