

Базовая математика для искусственного интеллекта

Математика нового уровня
для эффективных
и успешных систем ИИ



Хала Нельсон

Содержание

Отзывы о книге "Базовая математика для искусственного интеллекта"	15
Введение	17
Почему я написала эту книгу	17
Кому предназначена книга	19
Кому противопоказана книга	20
Как в книге представлена математика.....	20
Инфографика.....	22
Требования к математической подготовке для понимания материала книги	23
Краткий обзор глав.....	23
Мои любимые книги по искусственному интеллекту	28
Условные обозначения и соглашения	29
Примеры применения кода.....	30
Платформа онлайн-обучения O'Reilly	31
Как с нами связаться?	31
Благодарности.....	31
 ГЛАВА 1. Почему так важно изучать математику искусственного интеллекта?	33
Что такое искусственный интеллект	34
Секрет популярности искусственного интеллекта.....	35
Возможности искусственного интеллекта	36
Конкретные задачи агента искусственного интеллекта.....	36
Ограничения искусственного интеллекта	38
Что происходит, когда системы искусственного интеллекта дают сбой	41
Направления развития искусственного интеллекта	41
Кто сегодня вносит основной вклад в развитие искусственного интеллекта.....	43
Какая математика используется в искусственном интеллекте	43
Итоги и перспективы	44
 ГЛАВА 2. Данные, данные, данные	45
Данные для искусственного интеллекта	46
Реальные и имитационные данные.....	48
Математические модели: линейные и нелинейные	48
Пример реальных данных.....	50
Пример имитационных данных	53

Математические модели: симуляции и искусственный язык	56
Как получить данные	59
Словарь распределений данных, вероятностей, статистики	61
Случайные величины.....	62
Распределение вероятностей	62
Маргинальные (частные) вероятности.....	62
Равномерное и нормальное распределения	63
Условные вероятности и теорема Байеса	63
Условные вероятности и совместные распределения	63
Предварительное распределение, последующее распределение, функция правдоподобия	64
Смеси распределений	64
Суммы и произведения случайных величин	64
Использование графов для представления совместных распределений вероятностей.....	64
Ожидание, среднее значение, вариация, неопределенность.....	65
Ковариация и корреляция	65
Марковский процесс.....	66
Нормализация, масштабирование, стандартизация случайной переменной или набора данных.....	66
Общие примеры	67
Непрерывные и дискретные распределения (плотность vs масса)	67
Сила совместной функции плотности вероятности	69
Распределение данных: равномерное распределение	70
Распределение данных: колоколообразное нормальное (гауссово) распределение	72
Распределение данных: другие важные и часто используемые распределения.....	75
Варианты значений слова "распределение"	79
А/В-тестирование	80
Итоги и перспективы	81
ГЛАВА 3. Подгонка функций под данные	83
Лучшие традиционные модели машинного обучения	85
Численные и аналитические решения	87
Регрессия: предсказание числового значения	88
Обучающая функция	90
Функция потерь.....	92
Оптимизация	102
Логистическая регрессия: разделение на два класса	115
Обучающая функция	115
Функция потерь.....	116
Оптимизация	118
Регрессия softmax: классификация по нескольким категориям	118
Обучающая функция	120
Функция потерь.....	121
Оптимизация	122
Встраивание моделей в последний слой нейросети.....	123

Другие популярные методы машинного обучения и ансамбли методов	123
Машины опорных векторов	124
Деревья решений	128
Случайные леса	137
Кластеризация k средних	138
Показатели эффективности моделей классификации	138
Итоги и перспективы	140
ГЛАВА 4. Оптимизация нейронных сетей	143
Кора мозга и искусственные нейросети	143
Обучающая функция: полносвязные (плотные) нейросети с прямой связью	145
Нейросеть — это вычислительный граф, представляющий обучающую функцию	146
Линейная комбинация, добавление смещения и последующая активация	147
Основные функции активации	151
Универсальная аппроксимация функций	154
Теория аппроксимации в глубоком обучении	160
Функции потерь	160
Оптимизация	162
Математика и загадочный успех нейронных сетей	163
Градиентный спуск $\bar{\omega}^{i+1} = \bar{\omega}^i - \eta \nabla L(\bar{\omega}^i)$	164
Роль гиперпараметра скорости обучения η	166
Выпуклые и невыпуклые ландшафты	169
Стохастический градиентный спуск	172
Инициализация весов $\bar{\omega}_0$ в оптимизации	173
Методы регуляризации	174
Отсев	174
Ранняя остановка	175
Пакетная нормализация каждого слоя	175
Регулирование размера весов наложением штрафа на их норму	177
Наложение штрафа на нормы l^2 и l^1	179
Роль параметра α в регуляризации	180
Примеры гиперпараметров в машинном обучении	181
Цепное правило и обратное распространение ошибки: вычисление $\nabla L(\bar{\omega}^i)$	182
Метод обратного распространения ошибки практически не отличается от того, как обучается наш мозг	183
Почему лучше использовать обратное распространение ошибки	184
Обратное распространение ошибки в деталях	184
Оценка значимости признаков входных данных	186
Итоги и перспективы	186
ГЛАВА 5. Сверточные нейронные сети и компьютерное зрение	189
Свертка и кросс-корреляция	190
Инвариантность и эквивариантность переноса	194
Свертка в обычном пространстве — это произведение в частотном пространстве	195

Свертка с точки зрения проектирования системы	195
Свертка и импульсная характеристика линейных и инвариантных по переносу систем	197
Свертка и одномерные дискретные сигналы	199
Свертка и двумерные дискретные сигналы	200
Фильтрация изображений	202
Карта признаков	205
Нотация линейной алгебры	206
Одномерный случай: умножение на матрицу Тёплица	209
Двумерный случай: умножение на дважды блочно-циркулянтную матрицу	209
Пулинг	210
Сверточная нейросеть для классификации изображений	211
Итоги и перспективы	213

ГЛАВА 6. Сингулярное разложение: обработка изображений, обработка естественного языка и социальные сети	215
Факторизация матриц	216
Диагональные матрицы	219
Матрицы как линейные преобразования, действующие в пространстве	220
Воздействие A на правые сингулярные векторы	222
Воздействие A на стандартные единичные векторы и определяемый ими единичный квадрат	223
Воздействие A на единичную окружность	223
Преобразование окружности в эллипс методом сингулярного разложения	224
Матрицы поворота и отражения	225
Воздействие A на общий вектор $\bar{x}$	226
Три способа умножения матриц	227
Большая картина	228
Число обусловленности и вычислительная устойчивость	230
Компоненты сингулярного разложения	230
Сингулярное и спектральное разложения	231
Вычисление сингулярного разложения	233
Численное вычисление собственного вектора	233
Псевдоинверсия	235
Применение сингулярного разложения к изображениям	236
Метод главных компонент и снижение размерности	239
Метод главных компонент и кластеризация	241
Приложение для социальных сетей	241
Латентно-семантический анализ	242
Рандомизированное сингулярное разложение	243
Итоги и перспективы	243

ГЛАВА 7. Искусственный интеллект для естественного языка и финансов: векторизация и временные ряды	245
Искусственный интеллект в обработке естественного языка	248
Подготовка данных естественного языка к машинной обработке	249
Статистические модели и логарифмическая функция	252

Закон Ципфа для подсчета терминов	253
Векторные представления документов на естественном языке.....	253
Векторное представление частоты термина документа или мешка слов	253
Векторное представление частоты термина и обратной частоты термина документа.....	254
Векторное представление тематики документа, определенной с помощью скрытого семантического анализа	255
Тематическое векторное представление документа, определенное с помощью латентного размещения Дирихле.....	259
Тематическое векторное представление документа, определенное с помощью латентного дискриминантного анализа	260
Смысловые векторные представления слов и документов, определяемые встроенными нейросетями.....	261
Косинусное сходство	268
Приложения для обработки естественного языка	269
Анализ настроений	269
Фильтрация спама.....	270
Поиск и извлечение информации.....	271
Машинный перевод	273
Создание подписей к изображениям.....	273
Чат-боты.....	273
Другие приложения	274
Трансформеры и модели внимания	274
Архитектура трансформера	275
Механизм внимания	278
Трансформеры далеки от совершенства	282
Сверточные нейросети для данных временных рядов.....	282
Рекуррентные нейросети для данных временных рядов	284
Как работает рекуррентная нейросеть	285
Управляемые рекуррентные блоки и блоки долгой краткосрочной памяти.....	286
Пример данных естественного языка	287
Финансовый искусственный интеллект	287
Итоги и перспективы	288
ГЛАВА 8. Вероятностные генеративные модели.....	289
Ценность генеративных моделей.....	290
Типичная математика генеративных моделей.....	292
Переключение мозга с детерминированного на вероятностное мышление	294
Оценка максимального правдоподобия	296
Явные и неявные модели плотности	298
Явно прослеживаемая плотность: полностью видимые сети доверия.....	299
Пример: генерация изображений с помощью PixelCNN и машинного звука с помощью WaveNet.....	299
Явно прослеживаемая плотность: нелинейный анализ независимых компонент с изменением переменных.....	302
Явно непрослеживаемая плотность: аппроксимация вариационных автокодировщиков вариационными методами.....	303

Явно непрослеживаемая плотность: аппроксимация машины Больцмана с помощью цепи Маркова.....	305
Неявная плотность — цепь Маркова: генеративная стохастическая сеть	305
Неявная плотность — направление: генеративно-состязательные сети	306
Принцип работы генеративно-состязательной сети	307
Пример: машинное обучение и генеративные сети для физики высоких энергий	309
Другие генеративные модели	312
Наивная байесовская модель классификации	313
Модель гауссовой смеси	314
Эволюция генеративных моделей	315
Сети Хопфилда	317
Машина Больцмана	317
Ограниченная машина Больцмана (явная плотность и трудноразрешимость)	317
Исходный автокодировщик	319
Вероятностное языковое моделирование	319
Итоги и перспективы	322

ГЛАВА 9. Графовые модели..... 323

Графы: узлы, грани и их признаки.....	325
Пример: алгоритм PageRank.....	328
Инверсия матриц с помощью графов	333
Графы Кэли для групп: чистая алгебра и параллельные вычисления	333
Передача сообщений внутри графа	334
Безграничные возможности применения графов	335
Сети головного мозга	336
Распространение заболеваний	337
Распространение информации	337
Обнаружение и отслеживание распространения фейков	338
Рекомендательные системы веб-масштаба	339
Борьба с онкологией	340
Биохимические графы	341
Генерация молекулярных графов для выявления структуры лекарственных препаратов и белков	342
Сети цитирования	342
Социальные медиасети и прогнозирование общественного воздействия	342
Социологические структуры	343
Байесовские сети	343
Прогноз трафика	343
Логистика и исследование операций	344
Языковые модели	344
Графовая структура Всемирной паутины	346
Автоматический анализ компьютерных программ	347
Структуры данных в вычислительной технике	347
Балансировка нагрузки в распределенных сетях	348
Искусственные нейросети	349
Случайные блуждания по графам	350
Обучение представлению узлов	352

Задачи для графовых нейросетей.....	353
Классификация узлов	353
Классификация графов.....	354
Кластеризация и обнаружение сообществ.....	354
Генерация графов.....	355
Максимизация влияния	355
Предсказание ссылок.....	355
Динамические графовые модели	356
Байесовские сети	357
Байесовская сеть как компактная таблица условных вероятностей	359
Создание предсказаний при помощи байесовской сети	359
Байесовские сети — это не каузальные сети, а сети доверия.....	360
Что нужно знать о байесовских сетях.....	360
Цепи, вилки и коллаидеры	361
Как создать байесовскую сеть для всех переменных из набора данных?	362
Графические диаграммы вероятностного каузального моделирования	363
Краткая история теории графов	365
Основные положения теории графов	367
Остовные деревья и кратчайшие остовные деревья	367
Наборы разрезов и разрезанные вершины	367
Планарность	368
Графы как векторные пространства.....	368
Реализуемость	369
Раскраска и сопоставление	369
Перечисление	370
Алгоритмы и вычислительные аспекты графов	370
Итоги и перспективы	371
ГЛАВА 10. Исследование операций.....	373
Отсутствие бесплатного обеда.....	375
Анализ сложности и нотация $O()$	376
Оптимизация — сердце исследования операций	379
Рассуждения об оптимизации	383
Оптимизация: конечные размерности без ограничений	383
Оптимизация: конечные размерности, ограниченные множители Лагранжа.....	384
Оптимизация: бесконечные размерности, вариационное исчисление	386
Оптимизация в сетях	391
Задача о коммивояжере	391
Минимальное остовное дерево.....	392
Кратчайший путь	393
Максимальный поток и минимальный разрез.....	394
Максимальный поток и минимальная стоимость	395
Метод критического пути для разработки проекта	396
Задача об n ферзях.....	396
Линейная оптимизация	397
Общая и стандартная формы	398
Визуализация задачи линейной оптимизации в двух измерениях.....	399

От выпуклости к линейности.....	400
Геометрия линейной оптимизации.....	402
Симплекс-метод	404
Транспортная задача и задача о назначении	411
Двойственность, релаксация Лагранжа, теневые цены, максимин, минимакс и все такое.....	412
Чувствительность.....	425
Теория игр и мультиагентные системы.....	425
Очереди.....	427
Управление запасами	428
Машинное обучение в исследовании операций	428
Уравнение Гамильтона — Якоби — Беллмана	430
Исследование операций в искусственном интеллекте.....	430
Итоги и перспективы	431
ГЛАВА 11. Вероятность	433
Когда в книге встречается вероятность?	434
Что еще важно знать в области искусственного интеллекта?	437
Каузальное моделирование и исчисление <i>do</i>	437
Альтернативный метод: исчисление <i>do</i>	439
Парадоксы и интерпретация диаграмм	442
Парадокс Монти Холла	442
Парадокс Берксона.....	444
Парадокс Симпсона	444
Большие случайные матрицы.....	446
Примеры случайных векторов и случайных матриц	447
Основные положения теории случайных матриц	450
Ансамбли случайных матриц	451
Плотность собственных значений суммы двух больших случайных матриц.....	453
Базовая математика для больших случайных матриц	453
Стохастические процессы	454
Процесс Бернулли	456
Процесс Пуассона	456
Случайное блуждание	457
Процесс Винера, или броуновское движение	457
Мартингейл.....	458
Процесс Леви.....	459
Процесс ветвления	459
Цепь Маркова	459
Лемма Ито.....	460
Марковские процессы принятия решений и обучение с подкреплением.....	461
Примеры обучения с подкреплением	461
Обучение с подкреплением как марковский процесс принятия решений	462
Обучение с подкреплением в контексте оптимального управления и нелинейной динамики	464
Библиотека Python для обучения с подкреплением.....	464

Строгие теоретические основания	464
Какие события вероятны?	465
Возможен ли более широкий диапазон случайных величин?	466
Вероятностная тройка (пространство выборок, сигма-алгебра, мера вероятности)	467
В чем сложность?	468
Случайная величина, ожидание, интегрирование	468
Распределение случайной величины и теорема о замене переменной	470
Дальнейшие шаги в строгой теории вероятностей	470
Универсальная теорема для нейросетей	471
Итоги и перспективы	471
ГЛАВА 12. Математическая логика	473
Логические структуры	474
Пропозициональная логика	474
От нескольких аксиом к целостной теории	477
Кодирование логики внутри агента	478
Как сочетаются детерминированное и вероятностное машинное обучение?	478
Логика первого порядка	479
Отношения между кванторами существования и всеобщности	480
Вероятностная логика	482
Нечеткая логика	482
Темпоральная логика	483
Сравнение с естественным человеческим языком	484
Машины и сложные математические рассуждения	485
Итоги и перспективы	485
ГЛАВА 13. Искусственный интеллект и дифференциальные уравнения в частных производных	487
Что собой представляет дифференциальное уравнение в частных производных?	488
Моделирование с помощью дифференциальных уравнений	490
Модели различных масштабов	490
Параметры дифференциального уравнения в частных производных	491
Как изменение одного параметра в дифференциальном уравнении в частных производных может обернуться серьезными проблемами	492
Насколько полезен искусственный интеллект?	493
Численные решения обладают огромной ценностью	494
Непрерывные и дискретные функции	495
Дифференциальные уравнения в частных производных в моей докторской диссертации	497
Дискретизация и проклятие размерности	499
Метод конечных разностей	500
Метод конечных элементов	506
Вариационные или энергетические методы	511
Методы Монте-Карло	512
Немного статистической механики: замечательное основное уравнение	515
Решения как ожидания лежащих в основе случайных процессов	517

Преобразование дифференциального уравнения в частных производных	518
Преобразование Фурье	518
Преобразование Лапласа	520
Операторы решения	522
Пример использования уравнения теплопроводности	522
Пример использования уравнения Пуассона	524
Метод простой итерации	526
Искусственный интеллект применительно к дифференциальным уравнениям в частных производных	533
Глубокое обучение для нахождения значений физических параметров	533
Глубокое обучение для расчета сеток	534
Глубокое обучение для аппроксимации операторов решения дифференциальных уравнений в частных производных	536
Численные решения высокоразмерных дифференциальных уравнений	543
Моделирование природных явлений непосредственно из данных	545
Дифференциальное уравнение в частных производных Гамильтона — Якоби — Беллмана для динамического программирования	547
Дифференциальные уравнения в частных производных для искусственного интеллекта	552
Другие аспекты дифференциальных уравнений в частных производных	553
Итоги и перспективы	555

ГЛАВА 14. Искусственный интеллект, этика, математика, право

и политика	557
Хороший искусственный интеллект	559
Вопросы политики	561
Что может пойти не так?	562
От математики к вооружению	562
Боевые отравляющие вещества	564
Искусственный интеллект и политика	564
Нежелательные результаты генеративных моделей	565
Как это исправить?	566
Решение проблемы недостаточной представленности в обучающих данных	566
Устранение погрешностей в векторах слов	566
Решение проблемы конфиденциальности	567
Решение проблемы справедливости	568
Встраивание моральных аспектов в искусственный интеллект	569
Демократизация и доступность искусственного интеллекта для неспециалистов	570
Установка приоритета на высококачественные данные	570
Отличие предвзятости от дискриминации	572
Излишняя шумиха	572
Заключительные размышления	573
Предметный указатель	575
Об авторе	590
Об изображении на обложке	591