

# Оглавление

Предисловие .....	19
Благодарности .....	21
<b>Глава 1 Введение .....</b>	<b>23</b>
1.1. Обучение с учителем .....	24
1.1.1. Задачи регрессии и классификации .....	24
1.1.2. Входы модели .....	26
1.1.3. Модели машинного обучения .....	26
1.1.4. Глубокие нейронные сети .....	27
1.1.5. Структурированные выходы .....	28
1.2. Обучение без учителя .....	28
1.2.1. Генеративные модели .....	30
1.2.2. Скрытые переменные .....	30
1.2.3. Связь между обучением с учителем и без учителя .....	32
1.3. Обучение с подкреплением .....	14
1.3.1. Два примера .....	34
1.4. Этика .....	36
1.5. Структура книги .....	38
1.6. Другие книги .....	38
1.7. Как читать эту книгу .....	39
<b>Глава 2 Обучение с учителем .....</b>	<b>41</b>
2.1. Обзор обучения с учителем .....	41
2.2. Пример с линейной регрессией .....	42
2.2.1. Модель одномерной линейной регрессии .....	43
2.2.2. Потери .....	43
2.2.3. Обучение .....	46
2.2.4. Тестирование .....	46
2.3. Краткие итоги главы .....	47

Заметки .....	47
Задачи .....	48
<b>Глава 3 Неглубокие нейронные сети .....</b>	<b>49</b>
3.1. Пример нейронной сети .....	49
3.1.1. Интуитивное представление нейронной сети .....	50
3.1.2. Изображения нейронных сетей .....	51
3.2. Универсальная теорема аппроксимации .....	53
3.3. Входы и выходы со многими переменными .....	54
3.3.1. Визуализация выходов со многими переменными .....	54
3.3.2. Визуализация входов со многими переменными .....	55
3.4. Неглубокие нейронные сети: общий случай .....	56
3.5. Терминология .....	59
3.6. Краткие итоги главы .....	60
Заметки .....	60
Задачи .....	63
<b>Глава 4 Глубокие нейронные сети .....</b>	<b>66</b>
4.1. Композиция нейронной сети .....	66
4.2. От композиции сетей к глубоким сетям .....	69
4.3. Глубокие нейронные сети .....	70
4.3.1. Гиперпараметры .....	72
4.4. Матричная нотация .....	73
4.4.1. Общая формулировка .....	73
4.5. Неглубокие сети vs глубокие сети .....	74
4.5.1. Способность аппроксимировать различные функции .....	74
4.5.2. Количество линейных областей на параметр .....	75
4.5.3. Эффективность глубины .....	76
4.5.4. Большие структурированные входы .....	76
4.5.5. Обучение и обобщение .....	76
4.6. Краткие итоги главы .....	77
Заметки .....	77
Задачи .....	79
<b>Глава 5 Функции потерь .....</b>	<b>82</b>
5.1. Максимальное правдоподобие .....	82
5.1.1. Вычисление распределения выходов .....	83
5.1.2. Критерий максимального правдоподобия .....	83
5.1.3. Максимизация логарифмического правдоподобия .....	85
5.1.4. Минимизация отрицательного логарифмического правдоподобия .....	86
5.1.5. Применение модели .....	86
5.2. Рецепт построения функций потерь .....	86
5.3. Пример 1: регрессия с одной переменной .....	87
5.3.1. Функция минимальных квадратичных потерь .....	88

5.3.2. Применение модели .....	90
5.3.3. Оценка дисперсии .....	90
5.3.4. Гетероскедастическая регрессия .....	90
5.4. Пример 2: бинарная классификация .....	92
5.5. Пример 3: мультиклассовая классификация .....	94
5.5.1. Предсказывание других типов данных .....	95
5.6. Множественные выходы .....	96
5.7. Потери кросс-энтропии .....	97
5.8. Краткие итоги главы .....	99
Заметки .....	99
Задачи .....	101
<b>Глава 6 Обучение моделей .....</b>	<b>105</b>
6.1 Градиентный спуск .....	105
6.1.1. Пример с линейной регрессией .....	106
6.1.2. Пример с моделью Габор .....	107
6.1.3. Локальные минимумы и седловые точки .....	110
6.2. Стохастический градиентный спуск .....	111
6.2.1. Батчи (пакеты) и эпохи .....	112
6.2.2. Свойства стохастического градиентного спуска .....	114
6.3. Импульс (momentum) .....	114
6.3.1. Ускоренный импульс Нестерова .....	115
6.4. Adam .....	116
6.5. Обучение гиперпараметров алгоритма .....	119
6.6. Краткие итоги главы .....	119
Заметки .....	120
Задачи .....	123
<b>Глава 7 Градиенты и инициализация .....</b>	<b>126</b>
7.1. Определения задач .....	126
7.2. Вычисление производных .....	128
7.3. Простой пример со скалярными параметрами .....	130
7.4. Алгоритм обратного распространения .....	135
7.4.1. Краткое изложение алгоритма обратного распространения .....	136
7.4.2. Алгоритмическое дифференцирование .....	137
7.4.3. Расширение для произвольных вычислительных графов .....	138
7.5. Инициализация параметров .....	138
7.5.1. Инициализация для прямого прохода .....	139
7.5.2. Инициализация для обратного прохода .....	140
7.5.3. Инициализация как для прямого, так и для обратного проходов .....	141
7.6. Пример обучающего кода .....	142
7.7. Краткие итоги главы .....	142
Заметки .....	142
Задачи .....	146

<b>Глава 8 Измерение производительности</b> .....	150
8.1. Обучение простой модели .....	150
8.2. Источники ошибок .....	152
8.2.1. Шум, смещение и дисперсия .....	153
8.2.2. Математическая формулировка ошибок на тесте .....	155
8.3. Уменьшение ошибки .....	156
8.3.1. Уменьшение дисперсии .....	157
8.3.2. Уменьшение смещения .....	157
8.3.3. Баланс смещение-дисперсия .....	157
8.4. Двойной спуск .....	159
8.4.1. Объяснение .....	161
8.5. Выбор гиперпараметров .....	165
8.6. Краткие итоги главы .....	166
Заметки .....	166
Задачи .....	170
<b>Глава 9 Регуляризация</b> .....	172
9.1. Регуляризация в явной форме .....	172
9.1.1. Вероятностная интерпретация .....	173
9.1.2. L2-регуляризация .....	174
9.2. Неявная регуляризация .....	175
9.2.1. Неявная регуляризация в градиентном спуске .....	175
9.2.2. Неявная регуляризация в стохастическом градиентном спуске .....	177
9.3. Эвристики для улучшения характеристик .....	177
9.3.1. Ранняя остановка .....	178
9.3.2. Ансамблевые методы .....	180
9.3.3. Дропаут .....	181
9.3.4. Наложение шума .....	183
9.3.5. Байесовский вывод .....	184
9.3.6. Трансферное и многозадачное обучение .....	185
9.3.7. Самообучение .....	187
9.3.8. Аугментация .....	187
9.4. Краткие итоги главы .....	188
Заметки .....	189
Задачи .....	196
<b>Глава 10 Сверточные сети</b> .....	197
10.1. Инвариантность и эквивариантность .....	198
10.2. Сверточные сети для одномерных входов .....	199
10.2.1. Одномерная операция свертки .....	199
10.2.2. Padding .....	200
10.2.3. Шаг, размер ядра и расширение .....	200
10.2.4. Сверточные слои .....	200
10.2.5. Каналы .....	201
10.2.6. Сверточные сети и рецептивные поля .....	202
10.2.7. Пример: MNIST-1D .....	203

10.3. Сверточные сети для двумерных входов .....	206
10.4. Увеличение и уменьшение размерности изображений (downsampling and upsampling) .....	208
10.4.1. Downsampling .....	208
10.4.2. Upsampling .....	209
10.4.3. Изменение количества каналов .....	209
10.5. Приложения .....	210
10.5.1. Классификация изображений .....	211
10.5.2. Обнаружение объектов .....	214
10.5.3. Семантическая сегментация .....	215
10.6. Краткие итоги главы .....	216
Заметки .....	217
Задачи .....	222
<b>Глава 11 Остаточные сети .....</b>	<b>224</b>
11.1. Последовательная обработка .....	224
11.1.1. Ограничения последовательной обработки .....	225
11.2. Остаточные соединения и остаточные блоки .....	227
11.2.1. Порядок операций в остаточных блоках .....	229
11.3. Взрывные градиенты в остаточных сетях .....	230
11.4. Пакетная нормализация .....	231
11.4.1. Затраты и преимущества пакетной нормализации .....	232
11.5. Общая остаточная архитектура .....	233
11.5.1. ResNet .....	233
11.5.2. DenseNet .....	235
11.5.3. U-Net и сети в виде «песочных часов» .....	236
11.6. Почему сети с остаточными соединениями работают так хорошо? .....	239
11.7. Краткие итоги главы .....	240
Заметки .....	240
Задачи .....	245
<b>Глава 12 Трансформеры .....</b>	<b>247</b>
12.1. Обработка текстовых данных .....	247
12.2. Механизм самовнимания на базе скалярного произведения .....	248
12.2.1. Вычисление и взвешивание значений .....	249
12.2.2. Вычисление весов внимания .....	250
12.2.3. Краткая суммаризация самовнимания .....	252
12.2.4. Матричная форма .....	252
12.3. Дополнения механизма самовнимания .....	253
12.3.1. Позиционное кодирование .....	254
12.3.2. Масштабированное самовнимание .....	254
12.3.3. Многоголовое самовнимание .....	256
12.4. Трансформеры .....	257
12.5. Трансформеры для обработки естественного языка .....	257
12.5.1. Разбиение на токены .....	257
12.5.2. Векторные представления (embeddings) .....	257

12.5.3. Модель трансформера	259
12.6. Пример модели энкодера: BERT	259
12.6.1. Предобучение	260
12.6.2. Тонкая настройка	261
12.7. Пример модели декодера: GPT-3	262
12.7.1. Моделирование языка	262
12.7.2. Маскированное самовнимание	263
12.7.3. Генерация текста декодером	264
12.7.4. GPT-3 и few-shot обучение	265
12.8. Пример модели энкодер-декодер: машинный перевод	266
12.9. Трансформеры для длинных последовательностей	268
12.10. Трансформеры для изображений	269
12.10.1. ImageGPT	270
12.10.2. Визуальный трансформер (Vision Transformer — ViT)	271
12.10.3. Мультимасштабные визуальные трансформеры	272
12.11. Краткие итоги главы	273
Заметки	273
Задачи	282
<b>Глава 13 Графовые нейронные сети</b>	<b>284</b>
13.1. Что такое граф?	284
13.1.1. Типы графов	285
13.2. Представление графа	287
13.2.1. Свойства матрицы смежности	288
13.2.2. Перестановка индексов вершин	289
13.3. Графовые нейронные сети, задачи и функции потерь	290
13.3.1. Задачи и функции потерь	290
13.4. Сверточные графовые сети	292
13.4.1. Эквивариантность и инвариантность	293
13.4.2. Общие параметры	293
13.4.3. Пример GCN слоя	294
13.5. Пример: классификация графа	295
13.5.1. Обучение батчами	295
13.6. Индуктивные модели vs трансдуктивные модели	296
13.7. Пример: классификация вершин	297
13.7.1. Выбор батчей	298
13.8. Слои сверточных графовых сетей	300
13.8.1. Комбинирование текущей вершины и агрегированных соседей	301
13.8.2. Остаточные соединения	301
13.8.3. Агрегация средним	301
13.8.4. Нормализация Кипфа	302
13.8.5. Агрегация с max pooling	302
13.8.6. Агрегация с учетом внимания	302
13.9. Реберные графы	303
13.10. Краткие итоги главы	305

Заметки .....	306
Задачи .....	312
<b>Глава 14 Обучение без учителя .....</b>	<b>314</b>
14.1. Таксономия моделей обучения без учителя .....	314
14.2. Что нужно для хорошей генеративной модели? .....	316
14.3. Количественная оценка качества .....	317
14.4. Краткие итоги главы .....	320
Заметки .....	321
<b>Глава 15 Состязательные генеративные сети .....</b>	<b>322</b>
15.1. Дискриминация как сигнал .....	322
15.1.1. Функция потерь GAN .....	324
15.1.2. Обучение GAN .....	324
15.1.3. Глубокие сверточные GAN .....	326
15.1.4. Сложности обучения GAN .....	326
15.2. Улучшение стабильности .....	328
15.2.1. Анализ функции потерь GAN .....	328
15.2.2. Исчезающие градиенты .....	329
15.2.3. Расстояние Васерштейна .....	330
15.2.4. Расстояние Васерштейна для дискретных распределений .....	330
15.2.5. Расстояние Васерштейна для непрерывных распределений .....	332
15.2.6. Функция потерь Васерштейна для GAN .....	332
15.3. Прогрессивный рост, дискриминатор на минибатчах и усечение .....	333
15.4. Генерация с учетом условий .....	336
15.4.1. Conditional GAN .....	336
15.4.2. GAN со вспомогательной классификацией .....	336
15.4.3. InfoGAN .....	336
15.5. Трансляция изображений .....	339
15.5.1. Pix2Pix .....	339
15.5.2. Состязательные потери .....	341
15.5.3. CycleGAN .....	341
15.6. StyleGAN .....	342
15.7. Краткие итоги главы .....	346
Заметки .....	346
Задачи .....	351
<b>Глава 16 Нормализующие потоки .....</b>	<b>352</b>
16.1. 1D: пример .....	352
16.1.1. Оценка вероятности .....	352
16.1.2. Прямое и обратное отображения .....	353
16.1.3. Обучение .....	355
16.2. Общий случай .....	355
16.2.1. Прямое отображение с глубокой нейронной сетью .....	356
16.2.2. Желательные свойства сетевых слоев .....	357
16.3. Инвертируемые сетевые слои .....	358

16.3.1. Линейные потоки	358
16.3.2. Поэлементные потоки	359
16.3.3. Связывающие потоки	360
16.3.4. Авторегрессионные потоки	361
16.3.5. Инверсные авторегрессионные потоки	362
16.3.6. Остаточные потоки: iRevNet	363
16.3.7. Остаточные потоки и сжимающие отображения: iResNet	364
16.4. Мультимасштабные потоки	366
16.5. Приложения	367
16.5.1. Моделирование распределений	367
16.5.2. Синтез	368
16.5.3. Аппроксимация других моделей плотности	369
16.6. Краткие итоги главы	371
Заметки	371
Задачи	375
<b>Глава 17 Вариационные автокодировщики</b>	<b>377</b>
17.1. Модели латентных переменных	377
17.1.1. Пример: смесь гауссиан	378
17.2. Нелинейная модель латентных переменных	379
17.2.1. Генерация	380
17.3. Обучение	380
17.3.1. Нижняя вариационная граница (Evidence lower bound — ELBO)	381
17.3.2. Неравенство Йенсена	381
17.3.3. Получение оценки	382
17.4. Свойства ELBO	384
17.4.1. Точная граница	384
17.4.2. ELBO как потери реконструкции минус KL-расстояние до априорного распределения	385
17.5. Вариационная аппроксимация	386
17.6. Вариационный автокодировщик	387
17.6.1. Алгоритм VAE	388
17.7. Прием с репараметризацией	390
17.8. Приложения	390
17.8.1. Аппроксимация вероятности образца	391
17.8.2. Генерация	392
17.8.3. Ресинтез	393
17.8.4. Распутывание	394
17.9. Краткие итоги главы	394
Заметки	395
Задачи	399
<b>Глава 18 Диффузионные модели</b>	<b>401</b>
18.1. Обзор	401
18.2. Энкодер (прямой процесс)	402

18.2.1. Диффузионное ядро $q(\mathbf{z}_t \mathbf{x})$ .....	404
18.2.2. Маргинальные распределения $q(\mathbf{z}_t)$ .....	405
18.2.3. Условное распределение $q(\mathbf{z}_{t-1} \mathbf{z}_t)$ .....	406
18.2.4. Условное диффузионное распределение $q(\mathbf{z}_{t-1} \mathbf{z}_t, \mathbf{x})$ .....	406
18.3. Модель декодера (обратный процесс) .....	409
18.4. Обучение .....	409
18.4.1. Вариационная нижняя граница (ELBO) .....	410
18.4.2. Упрощение ELBO .....	410
18.4.3. Анализируя ELBO .....	412
18.4.4. Диффузионная функция потерь .....	412
18.4.5. Процедура обучения .....	413
18.5. Репараметризация функции потерь .....	414
18.5.1. Репараметризация целевой переменной .....	414
18.5.2. Репараметризация сети .....	415
18.6. Реализация .....	416
18.6.1. Приложение к изображениям .....	417
18.6.2. Улучшение скорости генерации .....	418
18.6.3. Условная генерация .....	419
18.6.4. Улучшение качества генерации .....	420
18.7. Краткие итоги главы .....	422
Заметки .....	423
Задачи .....	426

<b>Глава 19 Обучение с подкреплением</b> .....	<b>429</b>
19.1. Марковские процессы принятия решений, возвраты и стратегии .....	430
19.1.1. Марковский процесс .....	430
19.1.2. Марковский процесс вознаграждения .....	430
19.1.3. Марковский процесс принятия решений .....	431
19.1.4. Частично наблюдаемые марковские процессы принятия решений .....	432
19.1.5. Стратегия .....	433
19.2. Ожидаемый возврат .....	433
19.2.1. Ценность состояния и действия .....	434
19.2.2. Оптимальная стратегия .....	435
19.2.3. Уравнения Беллмана .....	436
19.3. Табличное обучение с подкреплением .....	438
19.3.1. Динамическое программирование .....	438
19.3.2. Методы Монте-Карло .....	440
19.3.3. Методы темпоральной разности .....	441
19.4. Q-обучение с подгонкой .....	442
19.4.1. Глубокие Q-сети для игры в игры ATARI .....	443
19.4.2. Двойное Q-обучение и двойные глубокие Q-сети .....	445
19.5. Методы градиента стратегий .....	446
19.5.1. Вывод обновлений градиента .....	446
19.5.2. Алгоритм REINFORCE .....	448
19.5.3. Базовые значения (бейзлайны) .....	449

19.5.4. Базовые значения, зависящие от состояний .....	450
19.6. Методы актер-критик .....	451
19.7. Офлайн-обучение с подкреплением .....	451
19.8. Краткие итоги главы .....	453
Заметки .....	454
Задачи .....	458
<b>Глава 20 Почему глубокое обучение работает? .....</b>	<b>461</b>
20.1. Аргументация против глубокого обучения .....	461
20.1.1. Обучение .....	461
20.1.2. Обобщение .....	462
20.1.3. Нерациональная эффективность глубокого обучения .....	462
20.2. Факторы, которые влияют на производительность обучения .....	463
20.2.1. Набор данных .....	463
20.2.2. Регуляризация .....	464
20.2.3. Стохастический обучающий алгоритм .....	464
20.2.4. Перепараметризация .....	464
20.2.5. Функции активации .....	466
20.2.6. Инициализация .....	466
20.2.7. Глубина сети .....	467
20.3. Свойства функций потерь .....	467
20.3.1. Множественные глобальные минимумы .....	467
20.3.2. Путь к минимуму .....	468
20.3.3. Связи между минимумами .....	469
20.3.4. Кривизна поверхности потерь .....	469
20.4. Факторы, которые определяют обобщающую способность .....	471
20.4.1. Обучающие алгоритмы .....	471
20.4.2. Плоские минимумы .....	472
20.4.3. Архитектура .....	473
20.4.4. Норма весов .....	473
20.4.5. Перепараметризация .....	474
20.4.6. Покидая многообразие данных .....	474
20.5. Нужно ли нам так много параметров? .....	475
20.5.1. Прунинг .....	475
20.5.2. Дистилляция знаний .....	476
20.5.3. Обсуждение .....	477
20.6. Должны ли сети быть глубокими? .....	478
20.6.1. Сложность моделируемой функции .....	478
20.6.2. Разрешимость обучения .....	479
20.6.3. Индуктивное смещение .....	479
20.7. Краткие итоги главы .....	480
Задачи .....	480
<b>Глава 21 Этика глубокого обучения .....</b>	<b>481</b>
21.1. Выравнивание ценностей .....	482
21.1.1. Предвзятость и справедливость .....	483

21.1.2. Искусственная моральная ответственность	486
21.1.3. Прозрачность и непрозрачность	487
21.1.4. Объяснимость и интерпретируемость	487
21.2. Намеренное использование не по назначению	488
21.2.1. Распознавание и анализ лиц	489
21.2.2. Милитаризация и политическое вмешательство	489
21.2.3. Мошеннические действия	490
21.2.4. Приватность данных	490
21.3. Другие социальные, этические и профессиональные вопросы	491
21.3.1. Интеллектуальная собственность	491
21.3.2. Предвзятость в пользу автоматизации и утрата моральных навыков	492
21.3.3. Влияние на окружающую среду	492
21.3.4. Занятость и общество	492
21.3.5. Концентрация власти	493
21.4. Пример исследования	493
21.5. Ценностно-нейтральный идеал науки	494
21.6. Ответственные ИИ-исследования как социальная дилемма	495
21.6.1. Научные коммуникации	496
21.6.2. Разнообразие и разнородность	496
21.7. Пути в будущее	497
21.8. Краткие итоги главы	498
Задачи	499

<b>Приложение А Нотация</b>	501
Скаляры, векторы, матрицы и тензоры	501
Переменные и параметры	501
Множества	501
Функции	502
Минимизация и максимизация	502
Распределения вероятностей	503
Асимптотическая нотация	503
Разное	503

<b>Приложение В Математика</b>	504
В.1 Функции	504
В.1.1 Константа Липшица	504
В.1.2 Выпуклость	505
В.1.3 Специальные функции	505
В.1.4 Формула Стирлинга	506
В.2 Биномиальные коэффициенты	506
В.2.1 Автокорреляция	507
В.3 Вектор, матрицы и тензоры	507
В.3.1 Транспонирование	507
В.3.2 Векторные и матричные нормы	507
В.3.3 Произведение матриц	508

В.3.4	Скалярное произведение векторов	508
В.3.5	Инверсия	508
В.3.6	Подпространства	509
В.3.7	Спектр матрицы	509
В.3.8	Определитель и след	510
В.4	Специальные типы матриц	510
В.4.1	Диагональные матрицы	510
В.4.2	Треугольные матрицы	511
В.4.3	Ортогональные матрицы	511
В.4.4	Перестановочные матрицы	511
В.4.5	Линейная алгебра	511
В.4.6	Линейные уравнения в матричной форме	512
В.5	Матричное исчисление	512
<b>Приложение С</b>	<b>Вероятность</b>	<b>513</b>
С.1	Случайные переменные и распределения вероятностей	513
С.1.1	Совместная вероятность	514
С.1.2	Маргинализация	514
С.1.3	Условное распределение и правдоподобие	515
С.1.4	Теорема Байеса	515
С.1.5	Независимость	516
С.2	Математическое ожидание	517
С.2.1	Правила преобразования математических ожиданий	517
С.2.2	Среднее, дисперсия и ковариация	519
С.2.3	Тождество дисперсии	519
С.2.4	Стандартизация	520
С.3	Нормальное распределение вероятностей	521
С.3.1	Одномерное нормальное распределение	521
С.3.2	Многомерное нормальное распределение	522
С.3.3	Произведение двух нормальных распределений	522
С.3.4	Изменение переменной	523
С.4	Выборки	524
С.4.1	Выборки из нормального распределения	524
С.4.2	Наследственное сэмплирование	524
С.5	Расстояния между распределениями вероятностей	525
С.5.1	Расстояние Кульбака — Лейблера	525
С.5.2	Дивергенция Йенсена — Шеннона	526
С.5.3	Расстояние Фреше	526
С.5.4	Расстояния между нормальными распределениями	526
<b>Библиография</b>		<b>529</b>
<b>Алфавитный указатель</b>		<b>587</b>

# Глава 1

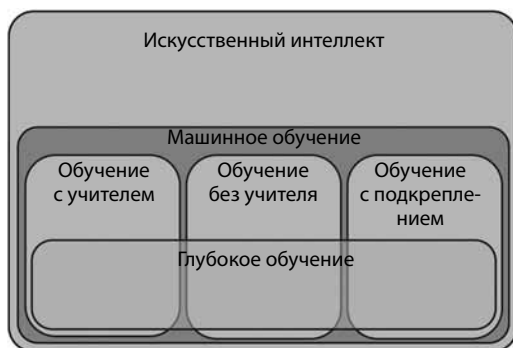
## Введение

*Искусственным интеллектом*, сокращенно ИИ (*англ.* AI), называют системы, имитирующие разумное поведение. Отрасль создания таких систем использует обширный диапазон подходов, базирующихся на логике, поиске и вероятностных рассуждениях. *Машинное обучение* — это составная часть ИИ, которая обучается принятию решений путем подгонки математических моделей к данным наблюдений. Эта область развивается бурными темпами и теперь уже (ошибочно) воспринимается как синоним термина ИИ.

*Глубокие нейронные сети* — один из типов моделей машинного обучения. Когда мы обучаем их на данных, речь идет о *глубоком обучении*. К моменту написания этой книги глубокие сети — это самая мощная и практичная из моделей машинного обучения, они часто встречаются в повседневной жизни. Стало привычным переводить тексты на иностранных языках, используя алгоритмы *обработки естественного языка*, искать в интернете изображения некоторого объекта, используя системы *компьютерного видения*, или беседовать с цифровым помощником с использованием *речевого интерфейса*. Все эти приложения оснащены движками глубокого обучения.

Как предполагает заголовок, цель этой книги — помочь незнакомым с этой областью читателям понять принципы глубокого обучения. Книга не слишком теоретическая (в ней нет доказательств теорем) и не то чтобы очень практическая (в ней почти нет образцов программного кода). В ней просто объясняются основополагающие *идеи*; освоив этот том, читатель сможет применить принципы глубокого обучения к новым ситуациям, не имеющим известного рецепта для достижения успеха.

Методы машинного обучения можно грубо разбить на три области: обучение с учителем, обучение без учителя и обучение с подкреплением. В момент написания этого текста самые передовые методы во всех трех областях основаны на глубоком обучении (рис. 1.1). Вводная глава содержит описание высокого уровня для всех трех областей, и эта таксономия примерно так же отражается в структуре книги. Нравится нам это или нет, глубокое обучение вот-вот изменит наш мир, и не всегда позитивно. Поэтому в последних главах содержатся краткие основы этики ИИ. Завершают труд советы, как извлечь из него наибольшую пользу.



**Рис. 1.1.** Машинное обучение — это область искусственного интеллекта, которая сопоставляет математические модели с наблюдаемыми данными. Она грубо подразделяется на обучение с учителем, обучение без учителя и обучение с подкреплением. Глубокие нейронные сети вносят вклад в каждую из этих областей

## 1.1. Обучение с учителем

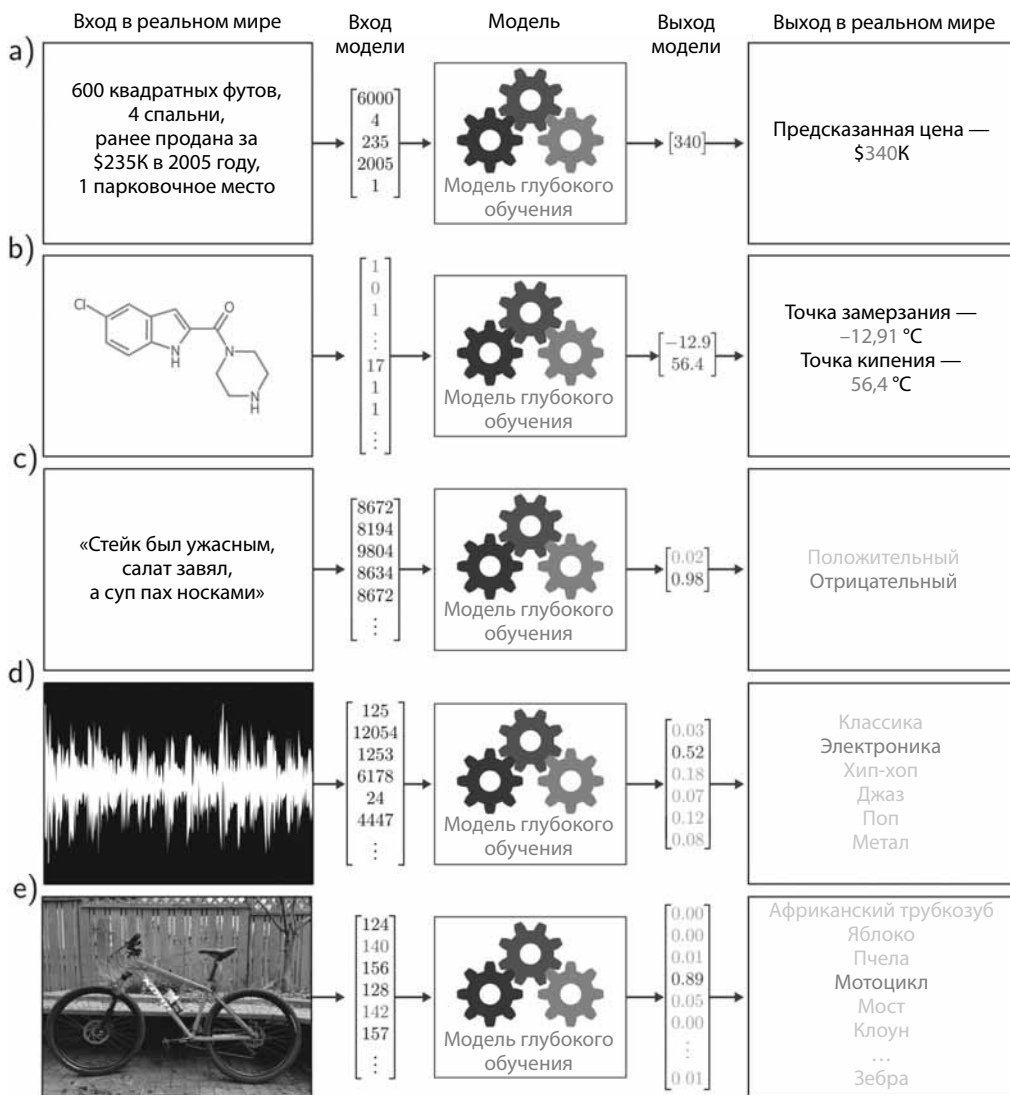
Модели обучения с учителем задают отображение данных на входе в предсказания на выходе. В следующих разделах мы обсудим входы, выходы, саму модель и само понятие «обучить» модель.

### 1.1.1. Задачи регрессии и классификации

На рис. 1.2 показаны несколько задач регрессии и классификации. В каждом случае на вход подается значимый в реальном мире объект (предложение, звуковой файл, изображение и т. д.), который кодируется как числовой вектор. Этот вектор образует вход модели. Модель отображает вход на выходной вектор, который затем «транслируется» обратно в значимое в реальном мире предсказание. Сначала мы рассмотрим только входы и выходы, считая модель черным ящиком, который поглощает один числовой вектор и возвращает другой числовой вектор.

Модель на рис. 1.2a предсказывает цену дома, основываясь на входных характеристиках, таких как площадь и количество спален. Это задача *регрессии*, поскольку модель возвращает числовое значение (а не отнесение к категории). По контрасту модель 1.2b берет в качестве входа химическую структуру молекулы и предсказывает точки кипения и плавления. Это задача *многомерной регрессии*, поскольку предсказывается более чем одно число.

Модель на рис. 1.2c получает на входе текстовую строку, содержащую отзыв о ресторане, и предсказывает, положительный это отзыв или отрицательный. Это задача *бинарной классификации*, поскольку модель пытается сопоставить входные данные одной из двух категорий. Выходной вектор содержит вероятности, что ввод относится к одной из категорий. Рис. 1.2d и 1.2e описывают задачу многоклассовой классификации. Здесь модель относит вход к одной из  $N > 2$  категорий. В первом случае вход — это аудиофайл, и модель предсказывает жанр записанной в нем музыки. Во втором случае вход — это изображение, а модель предсказывает, какой объект оно содержит. В обоих случаях модель возвращает вектор размера  $N$ , который содержит вероятности принадлежности к каждой из  $N$  категорий.



**Рис. 1.2.** Задачи регрессии и классификации. а) Эта *регрессионная* модель берет числовой вектор, который содержит характеристики объекта недвижимости, и предсказывает цену объекта. б) Эта модель *многомерной регрессии* берет структуру химической молекулы и предсказывает для нее точки плавления и кипения. в) Эта модель *бинарной классификации* берет отзыв о ресторане и классифицирует его как положительный или отрицательный. д) Эта задача *многоклассовой классификации* относит фрагмент аудио к одному из  $N$  жанров. е) Вторая задача *многоклассовой классификации* определяет изображение как содержащее один из  $N$  возможных объектов