
Оглавление

Предисловие	9
Введение	9
Условные обозначения	12
Примеры кода	13
Благодарности	13
Особая благодарность Элис Чжен	13
Особая благодарность Аманды Казари	14
Глава 1. Процесс машинного обучения	15
Данные	15
Задачи	15
Модели	16
Признаки	17
Оценка модели	18
Глава 2. Забавные трюки с простыми числами	19
Скаляры, векторы и пространства	21
Обработка счетчиков	22
Преобразование в двоичную форму	23
Квантование или разбиение на группы	25
Логарифмическое преобразование	31
Логарифмическое преобразование в действии	34
Степенное преобразование: обобщение логарифмического преобразования	40
Масштабирование признаков или нормализация	46
Масштабирование по минимуму	46
Стандартизация (масштабирование дисперсии)	47
Нормализация ℓ^2	49
Взаимодействие признаков	52
Отбор признаков	55
Выводы	57
Библиография	57
Глава 3. Текстовые данные: выравнивание, фильтрация и разбиение	58
Множество: преобразование естественного текста в плоский вектор	59
Множество слов	59

Множество n-грамм	62
Фильтрация и очистка признаков	65
Стоп-слова	65
Фильтрация на основе частотности	66
Выделение основы	69
Атомарное значение: от слов к n-граммам и фразам	70
Разбор и токенизация	70
Извлечение устойчивых выражений для обнаружения фраз	71
Выводы	80
Библиография	81
Глава 4. Эффекты масштабирования признаков:	
от множества слов к TF-IDF	82
TF-IDF: простое преобразование множества слов	82
Тестирование метода	84
Создание массива данных для классификации	86
Масштабирование множества слов с помощью преобразования TF-IDF	87
Классификация с логистической регрессией	88
Настройка логистической регрессии с помощью регуляризации	90
Глубокое погружение: что же происходит?	96
Выводы	99
Библиография	100
Глава 5. Категориальные переменные: подсчет яиц	
в эпоху роботизированных цыплят	101
Кодирование категориальных переменных	102
Кодирование одного состояния	102
Фиктивное кодирование	103
Эффективное кодирование	106
Плюсы и минусы кодирования категориальных переменных	108
Обработка больших категориальных переменных	108
Хеширование признаков	109
Подсчет бинов	113
Выводы	122
Библиография	123
Глава 6. Уменьшение размерности: сдавливание данных	
в блин с помощью PCA	125
Ключевая идея	125
Деривация	128
Линейная проекция	129

Дисперсия и эмпирическая дисперсия	130
Главные компоненты: первая формулировка	131
Основные компоненты: матрично-векторное представление	131
Общее решение основных компонент	132
Преобразование признаков	132
Реализация PCA	133
PCA в действии	134
Отбеливание и ZCA	136
Обсуждение метода PCA и его ограничений	137
Варианты использования	139
Выводы	141
Библиография	142
Глава 7. Нелинейное извлечение признаков с помощью стекирования	
моделей методом k -средних	143
Кластеризация с помощью метода k -средних	145
Кластеризация как разделение пространства	148
Применение метода k -средних для извлечения признаков	
задачи классификации	151
Альтернативное плотное извлечение признаков	158
Плюсы, минусы и подводные камни	158
Выводы	161
Библиография	162
Глава 8. Автоматизация извлечения признаков: извлечение	
признаков изображения и глубокое обучение	163
Простейшие признаки изображения (и почему они не работают)	164
Извлечение признаков вручную: алгоритмы SIFT и HOG	165
Градиенты изображения	166
Гистограмма направленных градиентов	170
Архитектура SIFT	174
Обучение признаков изображения с помощью нейронной	
сети глубокого обучения	175
Полностью связанные слои	176
Сверточные слои	177
Усеченное линейное преобразование (ReLU— Rectified Linear Unit)	182
Уровни нормализации ответов	183
Объединение слоев	185
Структура AlexNet	185
Выводы	189
Библиография	189

Глава 9. Назад в признаки: рекомендации научных статей	191
Совместная фильтрация на основе элементов	191
Первый проход: ввод данных, очистка и разбор признаков	193
Рекомендации научных статей: наивный подход	193
Второй проход: больше инженерных разработок и более разумная модель	201
Рекомендации научных статей: вторая попытка	201
Третий проход: дополнительные признаки = дополнительная информация	209
Рекомендации научных статей: попытка третья	209
Выводы	212
Библиография	213
Приложение А. Основы линейного моделирования и линейной алгебры ...	214
Обзор линейной классификации	214
Анатомия матрицы	217
От векторов к подпространствам	218
Сингулярное разложение (SVD — Singular Value Decomposition)	220
Четыре основных подпространства матрицы данных	222
Решение линейной системы	225
Библиография	228
Алфавитный указатель	228
Об авторах	233
Изображение на обложке	234