

Spis treści

1. Wprowadzenie	1
1.1. Kto powinien przeczytać tę książkę?	9
1.2. Historyczne trendy deep learningu	11
I Podstawy matematyki stosowanej i systemów uczących się	27
2. Algebra liniowa	29
2.1. Skalary, wektory, macierze i tensory	29
2.2. Mnożenie macierzy i wektorów	32
2.3. Macierze jednostkowe i odwrotne	34
2.4. Zależność liniowa i zakres	35
2.5. Normy	37
2.6. Macierze i wektory specjalne	38
2.7. Rozkład na wartości własne	40
2.8. Dekompozycja wartości osobliwej	42
2.9. Uogólniona macierz odwrotna (Moore'a-Penrose'a)	43
2.10. Operator śladowy	44
2.11. Wyznacznik	45
2.12. Przykład: analiza głównych składowych	45
3. Prawdopodobieństwo i teoria informacji	51
3.1. Dlaczego prawdopodobieństwo?	52
3.2. Zmienne losowe	54
3.3. Rozkłady prawdopodobieństwa	54
3.4. Prawdopodobieństwo brzegowe	56
3.5. Prawdopodobieństwo warunkowe	57
3.6. Reguła łańcuchowa w prawdopodobieństwie warunkowym	57
3.7. Niezależność oraz niezależność warunkowa	58
3.8. Wartość oczekiwana, wariancja i kowariancja	58

3.9.	Znane rozkłady prawdopodobieństwa	60
3.10.	Użyteczne cechy elementarnych funkcji	65
3.11.	Prawo Bayesa	68
3.12.	Techniczne szczegóły zmiennych ciągłych	68
3.13.	Teoria informacji	70
3.14.	Strukturalne modele probabilistyczne	73
4.	Obliczenia numeryczne	77
4.1.	Nadmiar i niedomiar	77
4.2.	Złe uwarunkowania	79
4.3.	Optymalizacja gradientowa	79
4.4.	Optymalizacja z ograniczeniami	89
4.5.	Przykład: liniowa metoda najmniejszych kwadratów	92
5.	Podstawy systemów uczących się	95
5.1.	Algorytmy uczenia się	96
5.2.	Pojemność, nadmierne dopasowanie i niedopasowanie	108
5.3.	Hiperparametry i zbiory walidacyjne	118
5.4.	Estymatory, obciążenie i wariancja	120
5.5.	Metoda maksymalnej wiarygodności	129
5.6.	Statystyki Bayesa	133
5.7.	Algorytmy nadzorowanego uczenia się	138
5.8.	Algorytmy nienadzorowanego uczenia się	143
5.9.	Metoda gradientu stochastycznego	150
5.10.	Tworzenie algorytmu dla systemu uczącego się	152
5.11.	Wyzwania motywujące deep learning	153
II	Głębokie sieci: nowoczesne praktyki	163
6.	Głębokie sieci jednokierunkowe	165
6.1.	Przykład: uczenie się funkcji XOR	168
6.2.	Uczenie się oparte na gradiencie	173
6.3.	Jednostki ukryte	188
6.4.	Projekt architektury	195
6.5.	Propagacja wsteczna i inne algorytmy różniczkowania	201
6.6.	Uwagi historyczne	221
7.	Regularyzacja w deep learningu	225
7.1.	Standardowe kary dla parametrów	227
7.2.	Standardowe kary jako optymalizacja z ograniczeniami	234

7.3.	Regularyzacja i problemy niedoograniczone	236
7.4.	Powiększanie zbioru danych	237
7.5.	Odporność na szum	239
7.6.	Uczenie się częściowo nadzorowane	241
7.7.	Uczenie się wielozadaniowe	242
7.8.	Wczesne zatrzymanie	243
7.9.	Wiązanie i współdzielenie parametrów	250
7.10.	Rzadko wypełnione reprezentacje	252
7.11.	Bagging i inne metody zespołowe	254
7.12.	Odrzucanie	256
7.13.	Szkolenie antagonistyczne	266
7.14.	Odległość styczna, propagacja stycznej oraz klasyfikator stycznej do różnorodności	268
8.	Optymalizacja w celu szkolenia głębokich modeli	273
8.1.	Czym uczenie się różni się od czystej optymalizacji	274
8.2.	Wyzwania związane z optymalizacją sieci neuronowej	281
8.3.	Podstawowe algorytmy	293
8.4.	Strategie nadawania parametrom wartości początkowych	299
8.5.	Algorytmy z adaptacyjną szybkością uczenia się	306
8.6.	Aproksymacyjne metody drugiego rzędu	310
8.7.	Strategie optymalizacji i meta-algorytmy	317
9.	Sieci spłotowe	331
9.1.	Splot jako działanie	332
9.2.	Uzasadnienie	334
9.3.	Redukcja	340
9.4.	Splot i redukcja jako nieskończenie silny rozkład aprioryczny	346
9.5.	Warianty podstawowej funkcji spłotowej	347
9.6.	Strukturalne wyjścia	358
9.7.	Typy danych	359
9.8.	Efektywne algorytmy spłotu	361
9.9.	Cechy losowe lub nienadzorowane	362
9.10.	Neuronaukowe podstawy sieci spłotowych	364
9.11.	Sieci spłotowe a historia deep learningu	371
10.	Modelowanie sekwencyjne: sieci rekurencyjne i rekursywne	373
10.1.	Rozwijanie grafów obliczeniowych	375
10.2.	Rekurencyjne sieci neuronowe	378

10.3.	Dwukierunkowe rekurencyjne sieci neuronowe	393
10.4.	Architektury koder-dekoder i sekwencja do sekwencji	394
10.5.	Głębokie sieci rekurencyjne	397
10.6.	Rekursywne sieci neuronowe	399
10.7.	Problem z zależnościami długoterminowymi	400
10.8.	Sieci stanu echa	403
10.9.	Nieszczelne jednostki i inne strategie dla wielu skali czasowych	406
10.10.	Długa pamięć krótkoterminowa i inne bramkowane sieci RNN	408
10.11.	Optymalizacja zależności długoterminowych	412
10.12.	Pamięć jawna	416
11.	Metodologia praktyczna	421
11.1.	Metryki wydajności	422
11.2.	Modele domyślnej linii bazowej	425
11.3.	Decyzja, czy zbierać więcej danych	426
11.4.	Wybór hiperparametrów	428
11.5.	Strategie debugowania	437
11.6.	Przykład: rozpoznawanie liczb wielocyfrowych	441
12.	Zastosowania	445
12.1.	Deep learning wielkoskalowy	445
12.2.	Rozpoznawanie obrazów	455
12.3.	Rozpoznawanie mowy	461
12.4.	Przetwarzanie języka naturalnego	464
12.5.	Inne zastosowania	482
III	Badania na polu deep learningu	491
13.	Liniowe modele czynnikowe	495
13.1.	Probabilistyczna analiza PCA i analiza czynnikowa	496
13.2.	Analiza składowych niezależnych (ICA)	497
13.3.	Powolna analiza cech	500
13.4.	Rzadkie kodowanie	502
13.5.	Poznanie różnorodności w analizie PCA	506
14.	Autokodery	509
14.1.	Autokodery niekompletne	510
14.2.	Autokodery z regularyzacją	511

14.3.	Reprezentacyjna potęga, rozmiar warstwy i głębokość	515
14.4.	Stochastyczne kodery i dekodery	516
14.5.	Autokodery z odsumianiem	517
14.6.	Poznanwanie rozmaitości z użyciem autokoderów	522
14.7.	Autokodery kurczliwe	527
14.8.	Predykcyjna rzadka dekompozycja	530
14.9.	Zastosowania autokoderów	531
15.	Poznanwanie reprezentacji	533
15.1.	Zachłanne nienadzorowane szkolenie wstępne warstwa po warstwie	535
15.2.	Transfer poznawania i adaptacja dziedziny	544
15.3.	Częściowo nadzorowane oswabadzanie czynników przyczynowych	548
15.4.	Reprezentacja rozproszona	554
15.5.	Wykładnicze zyski z głębokości	560
15.6.	Wskazówki do wykrywania przyczyn podstawowych	562
16.	Strukturalne modele probabilistyczne deep learningu	567
16.1.	Trudności w modelowaniu niestukturalnym	568
16.2.	Używanie grafów do opisu struktury modelu	572
16.3.	Próbkowanie z modeli graficznych	589
16.4.	Zalety modelowania strukturalnego	591
16.5.	Poznanwanie zależności	591
16.6.	Wnioskowanie i wnioskowanie przybliżone	592
16.7.	Strukturalne modele probabilistyczne w ujęciu deep learningu	594
17.	Metody Monte Carlo	599
17.1.	Próbkowanie i metody Monte Carlo	599
17.2.	Próbkowanie istotnościowe	601
17.3.	Metody Monte Carlo z łańcuchem Markowa	604
17.4.	Próbkowanie Gibbsa	608
17.5.	Problem mieszania między odseparowanymi trybami	609
18.	Zmagania z funkcją podziału	615
18.1.	Gradient wiarygodności logarytmicznej	616
18.2.	Stochastyczna maksymalna wiarygodność i kontrastywna dywergencja	617
18.3.	Pseudowiarygodność	625
18.4.	Dopasowywanie oceny i stosunku	628

18.5.	Dopasowywanie ocen z odszumianiem	630
18.6.	Estymacja kontrastywna szumu	630
18.7.	Szacowanie funkcji podziału	633
19.	Wnioskowanie przybliżone	641
19.1.	Wnioskowanie jako optymalizacja	642
19.2.	Maksymalizacja oczekiwania	644
19.3.	Wnioskowanie MAP i rzadkie kodowanie	645
19.4.	Wariacyjne wnioskowanie i uczenie się	648
19.5.	Poznanwanie wnioskowania przybliżonego	661
20.	Głębokie modele generatywne	665
20.1.	Maszyny Boltzmanna	665
20.2.	Ograniczone maszyny Boltzmanna	667
20.3.	Głębokie sieci przekonań	671
20.4.	Głębokie maszyny Boltzmanna	674
20.5.	Maszyny Boltzmanna dla danych rzeczywistych	688
20.6.	Splotowe maszyny Boltzmanna	695
20.7.	Maszyny Boltzmanna dla strukturalnych lub sekwencyjnych wartości wynikowych	697
20.8.	Inne maszyny Boltzmanna	698
20.9.	Propagacja wsteczna przez losowe działania	700
20.10.	Skierowane sieci generatywne	704
20.11.	Pobieranie próbek z autokoderów	724
20.12.	Generatywne sieci stochastyczne	727
20.13.	Inne schematy generowania	729
20.14.	Szacowanie modeli generatywnych	730
20.15.	Konkluzja	733
Bibliografia		735
Skorowidz		800