

Spis treści

Przedmowa 15

CZĘŚĆ I. PODSTAWY UCZENIA MASZYNOWEGO 25

1. Krajobraz uczenia maszynowego 27

Czym jest uczenie maszynowe? 28

Dlaczego warto korzystać z uczenia maszynowego? 28

Przykładowe zastosowania 31

Rodzaje systemów uczenia maszynowego 33

Uczenie nadzorowane i uczenie nienadzorowane 34

Uczenie wsadowe i uczenie przyrostowe 40

Uczenie z przykładów i uczenie z modelu 43

Główne problemy uczenia maszynowego 48

Niedobór danych uczących 50

Niereprezentatywne dane uczące 50

Dane kiepskiej jakości 51

Nieistotne cechy 52

Przetrenowanie danych uczących 52

Niedotrenowanie danych uczących 54

Podsumowanie 54

Testowanie i ocenianie 55

Strojenie hiperparametrów i dobór modelu 55

Niezgodność danych 56

Ćwiczenia 57

2. Nasz pierwszy projekt uczenia maszynowego 59

Praca z rzeczywistymi danymi 59

Przeanalizuj całokształt projektu 61

Określ zakres problemu 61

Wybierz metrykę wydajności 63

Sprawdź założenia 65

Zdobądź dane 65

Stwórz przestrzeń roboczą 66

Pobierz dane 68

Rzut oka na strukturę danych 70

Stwórz zbiór testowy 74

Odkrywaj i wizualizuj dane, aby zdobywać nowe informacje 78

Wizualizowanie danych geograficznych 78

Poszukiwanie korelacji 80

Eksperymentowanie z kombinacjami atrybutów 83

Przygotuj dane pod algorytmy uczenia maszynowego 84

Oczyszczanie danych 84

Obsługa tekstu i atrybutów kategorialnych 87

Niestandardowe transformatory 89

Skalowanie cech 90

Potoki transformujące 90

Wybór i uczenie modelu 92

Trenowanie i ocena modelu za pomocą zbioru uczącego 92

Dokładniejsze ocenianie za pomocą sprawdzianu krzyżowego 93

Wyreguluj swój model 96

Metoda przeszukiwania siatki 96

Metoda losowego przeszukiwania 98

Metody zespołowe 98

Analizuj najlepsze modele i ich błędy 98

Oceń system za pomocą zbioru testowego 99

Uruchom, monitoruj i utrzymuj swój system 100

Teraz Twoja kolej! 103

Ćwiczenia 103

3. Klasyfikacja 105

Zbiór danych MNIST 105

Uczenie klasyfikatora binarnego 107

Miary wydajności 108

Pomiar dokładności za pomocą sprawdzianu krzyżowego 108

Macierz pomyłek 110

Precyzja i pełność 111

Kompromis pomiędzy precyzją a pełnością 112

Wykres krzywej ROC 116

Klasyfikacja wieloklasowa 119

Analiza błędów 121

Klasyfikacja wieloetykietowa 124

Klasyfikacja wielowyjściowa 125

Ćwiczenia 127

4. Uczenie modeli 129

Regresja liniowa 130

Równanie normalne 131

Złożoność obliczeniowa 134

Gradient prosty 135

Wsadowy gradient prosty 138

Stochastyczny spadek wzdłuż gradientu 141

Schodzenie po gradiencie z minigrupami 143

Regresja wielomianowa 145

Krzywe uczenia 146

Regularyzowane modele liniowe 150

Regresja grzbietowa 150

Regresja metodą LASSO 153

Metoda elastycznej siatki 155

Wczesne zatrzymywanie 156

Regresja logistyczna 157

Szacowanie prawdopodobieństwa 158

Funkcje ucząca i kosztu 159

Granice decyzyjne 160

Regresja softmax 162

Ćwiczenia 166

5. Maszyny wektorów nośnych 167

Liniowa klasyfikacja SVM 167

Klasyfikacja miękkiego marginesu 168

Nieliniowa klasyfikacja SVM 170

Jądro wielomianowe 171

Cechy podobieństwa 172

Gaussowskie jądro RBF 173

Złożoność obliczeniowa 175

Regresja SVM 175

Mechanizm działania 177

Funkcja decyzyjna i prognozy 177

Cel uczenia 178

Programowanie kwadratowe 180

Problem dualny 181

Kernelizowane maszyny SVM 182

Przyrostowe maszyny SVM 185

Ćwiczenia 186

6. Drzewa decyzyjne 187

Uczenie i wizualizowanie drzewa decyzyjnego 187

Wyliczanie prognoz 188

Szacowanie prawdopodobieństw przynależności do klas 190

Algorytm uczący CART 191

Złożoność obliczeniowa 192

Wskaźnik Giniego czy entropia? 192

Hiperparametry regularyzacyjne 193

Regresja 194

Niestabilność 196

Ćwiczenia 197

7. Uczenie zespołowe i losowe lasy 199

Klasyfikatory głosujące 199

Agregacja i wklejanie 202

Agregacja i wklejanie w module Scikit-Learn 203

Ocena OOB 205

Rejony losowe i podprzestrzenie losowe 206

Losowe lasy 206

Zespół Extra-Trees 207

Istotność cech 207

Wzmacnianie 209

AdaBoost 209

Wzmacnianie gradientowe 212

Kontaminacja 217

Ćwiczenia 219

8. Redukcja wymiarowości 223

Klątwa wymiarowości 224

Główne strategie redukcji wymiarowości 225

Rzutowanie 225

Uczenie różnorodności 227

Analiza PCA 228

Zachowanie wariancji 229

Główne składowe 230

Rzutowanie na d wymiarów 231

Implementacja w module Scikit-Learn 232

Współczynnik wariancji wyjaśnionej 232

Wybór właściwej liczby wymiarów 232

Algorytm PCA w zastosowaniach kompresji 233

Losowa analiza PCA 234

Przyrostowa analiza PCA 235

Jądrowa analiza PCA 236

Wybór jądra i strojenie hiperparametrów 236

Algorytm LLE 239

Inne techniki redukcji wymiarowości 241

Ćwiczenia 241

9. Techniki uczenia nienadzorowanego 243

Analiza skupień 244

Algorytm centroidów 246

Granice algorytmu centroidów 255

Analiza skupień w segmentacji obrazu 256

Analiza skupień w przetwarzaniu wstępnym 257

Analiza skupień w uczeniu półnadzorowanym 259

Algorytm DBSCAN 262

Inne algorytmy analizy skupień 265

Mieszanki gaussowskie 266

Wykrywanie anomalii za pomocą mieszanin gaussowskich 271

Wyznaczanie liczby skupień 273

Modele bayesowskie mieszanin gaussowskich 275

Inne algorytmy służące do wykrywania anomalii i nowości 279

Ćwiczenia 280

CZĘŚĆ II. SIECI NEURONOWE I UCZENIE GŁĘBOKIE 283

10. Wprowadzenie do sztucznych sieci neuronowych i ich implementacji z użyciem interfejsu Keras 285

Od biologicznych do sztucznych neuronów 286

Neurony biologiczne 287

Operacje logiczne przy użyciu neuronów 288

Perceptron 289

Perceptron wielowarstwowy i propagacja wsteczna 293

Regresyjne perceptrony wielowarstwowe 297

Klasyfikacyjne perceptrony wielowarstwowe 298

Implementowanie perceptronów wielowarstwowych za pomocą interfejsu Keras 300

Instalacja modułu TensorFlow 2 301

Tworzenie klasyfikatora obrazów za pomocą interfejsu sekwencyjnego 302

Tworzenie regresyjnego perceptronu wielowarstwowego za pomocą interfejsu sekwencyjnego 311

Tworzenie złożonych modeli za pomocą interfejsu funkcyjnego 312

Tworzenie modeli dynamicznych za pomocą interfejsu podklasowego 316

Zapisywanie i odczytywanie modelu 318

Stosowanie wywołań zwrotnych 318

Wizualizacja danych za pomocą narzędzia TensorBoard 320

Dostrajanie hiperparametrów sieci neuronowej 323

Liczba warstw ukrytych 326

Liczba neuronów w poszczególnych warstwach ukrytych 327

Współczynnik uczenia, rozmiar grupy i pozostałe hiperparametry 328

Ćwiczenia 330

11. Uczenie głębokich sieci neuronowych 333

Problemy zanikających/eksplodujących gradientów 334

Inicjalizacje wag Glorota i He 334

Nienasycające funkcje aktywacji 336

Normalizacja wsadowa 340

Obcinanie gradientu 346

Wielokrotne stosowanie gotowych warstw 347

Uczenie transferowe w interfejsie Keras 348

Nienadzorowane uczenie wstępne 350

Uczenie wstępne za pomocą dodatkowego zadania 350

Szybsze optymalizatory 352

Optymalizacja momentum 352

Przyspieszony spadek wzdłuż gradientu (algorytm Nesterova) 353

AdaGrad 355

RMSProp 356

Optymalizatory Adam i Nadam 357

Harmonogramowanie współczynnika uczenia 359

Regularyzacja jako sposób zapobiegania przetrenowaniu 364

Regularyzacja l1 i l2 364

Porzucanie 365

Regularyzacja typu Monte Carlo (MC) 368

Regularyzacja typu max-norm 370

Podsumowanie i praktyczne wskazówki 371

Ćwiczenia 372

12. Modele niestandardowe i uczenie za pomocą modułu TensorFlow 375

Krótkie omówienie modułu TensorFlow 375

Korzystanie z modułu TensorFlow jak z biblioteki NumPy 379

Tensory i operacje 379

Tensory a biblioteka NumPy 381

Konwersje typów 381

Zmienne 381

Inne struktury danych 382

Dostosowywanie modeli i algorytmów uczenia 383

Niestandardowe funkcje straty 383

Zapisywanie i wczytywanie modeli zawierających elementy niestandardowe 384

Niestandardowe funkcje aktywacji, inicjalizatory, regularyzatory i ograniczenia 386

Niestandardowe wskaźniki 387

Niestandardowe warstwy 389

Niestandardowe modele 392

Funkcje straty i wskaźniki oparte na elementach wewnętrznych modelu 394

Obliczanie gradientów za pomocą różniczkowania automatycznego 396

Niestandardowe pętle uczenia 399

Funkcje i grafy modułu TensorFlow 402

AutoGraph i kreślenie 404

Reguły związane z funkcją TF 405

Ćwiczenia 406

13. Wczytywanie i wstępne przetwarzanie danych za pomocą modułu TensorFlow 409

Interfejs danych 410

Łączenie przekształceń 410

Tasowanie danych 412

Wstępne przetwarzanie danych 415

Składanie wszystkiego w całość 416

Pobieranie wstępne 417

Stosowanie zestawu danych z interfejsem tf.keras 418

Format TFRecord 419

Skompresowane pliki TFRecord 420

Wprowadzenie do buforów protokołów 420

Bufory protokołów w module TensorFlow 422

Wczytywanie i analizowanie składni obiektów Example 423

Obsługa list list za pomocą bufora protokołów SequenceExample 424

Wstępne przetwarzanie cech wejściowych 425

Kodowanie cech kategoryalnych za pomocą wektorów gorącojedynkowych 426

Kodowanie cech kategoryalnych za pomocą wektorów właściwościowych 428

Warstwy przetwarzania wstępnego w interfejsie Keras 431

TF Transform 433

Projekt TensorFlow Datasets (TFDS) 435

Ćwiczenia 436

14. Głębokie widzenie komputerowe za pomocą splotowych sieci neuronowych 439

Struktura kory wzrokowej 440

Warstwy splotowe 441

Filtry 443

Stosy map cech 444

Implementacja w module TensorFlow 446

Zużycie pamięci operacyjnej 448

Warstwa łącząca 449

Implementacja w module TensorFlow 451

Architektury splotowych sieci neuronowych 452

LeNet-5 454

AlexNet 455

GoogLeNet 458

VGGNet 461

ResNet 461

Xception 465

SENet 466

Implementacja sieci ResNet-34 za pomocą interfejsu Keras 468

Korzystanie z gotowych modeli w interfejsie Keras 469

Gotowe modele w uczeniu transferowym 471

Klasyfikowanie i lokalizowanie 473

Wykrywanie obiektów 474

W pełni połączone sieci spłotowe 476

Sieć YOLO 478

Segmentacja semantyczna 481

Ćwiczenia 484

15. Przetwarzanie sekwencji za pomocą sieci rekurencyjnych i spłotowych 487

Neurony i warstwy rekurencyjne 488

Komórki pamięci 490

Sekwencje wejść i wyjść 491

Uczenie sieci rekurencyjnych 492

Prognozowanie szeregów czasowych 493

Wskaźniki bazowe 494

Implementacja prostej sieci rekurencyjnej 494

Głębokie sieci rekurencyjne 496

Prognozowanie kilka taktów w przód 497

Obsługa długich sekwencji 500

Zwalczanie problemu niestabilnych gradientów 501

Zwalczanie problemu pamięci krótkotrwałej 503

Ćwiczenia 511

16. Przetwarzanie języka naturalnego za pomocą sieci rekurencyjnych i mechanizmów uwagi 513

Generowanie tekstów szekspirowskich za pomocą znakowej sieci rekurencyjnej 514

Tworzenie zestawu danych uczących 515

Rozdzielanie zestawu danych sekwencyjnych 515

Dzielenie zestawu danych sekwencyjnych na wiele ramek 516

Budowanie i uczenie modelu Char-RNN 518

Korzystanie z modelu Char-RNN 519

Generowanie sztucznego tekstu szekspirowskiego 519

Stanowe sieci rekurencyjne 520

Analiza sentymentów 522

Maskowanie 526

Korzystanie z gotowych reprezentacji właściwościowych 527

Sieć typu koder - dekodek służąca do neuronowego tłumaczenia maszynowego 529

Dwukierunkowe warstwy rekurencyjne 532

Przeszukiwanie wiązkowe 533

Mechanizmy uwagi 534

Mechanizm uwagi wizualnej 537

Liczy się tylko uwaga, czyli architektura transformatora 539

Współczesne innowacje w modelach językowych 546

Ćwiczenia 548

17. Uczenie reprezentacji za pomocą autokoderów i generatywnych sieci przeciwstawnych 551

Efektywne reprezentacje danych 552

Analiza PCA za pomocą niedopełnionego autokodera liniowego 554

Autokodery stosowe 555

Implementacja autokodera stosowego za pomocą interfejsu Keras 556

Wizualizowanie rekonstrukcji 557

Wizualizowanie zestawu danych Fashion MNIST 558

Nienadzorowane uczenie wstępne za pomocą autokoderów stosowych 558

Wiązanie wag 560

Uczenie autokoderów pojedynczo 561

Autokodery splotowe 562

Autokodery rekurencyjne 563

Autokodery odszumiające 564

Autokodery rzadkie 566

Autokodery wariacyjne 569

Generowanie obrazów Fashion MNIST 572

Generatywne sieci przeciwstawne 574

Problemy związane z uczeniem sieci GAN 577

Głębokie splotowe sieci GAN 579

Rozrost progresywny sieci GAN 582

Sieci StyleGAN 585

Ćwiczenia 587

18. Uczenie przez wzmacnianie 589

Uczenie się optymalizowania nagród 590

Wyszukiwanie strategii 591

Wprowadzenie do narzędzia OpenAI Gym 593

Sieci neuronowe jako strategię 597

Ocenianie czynności: problem przypisania zasługi 598

Gradienty strategii 600

Procesy decyzyjne Markowa 604

Uczenie metodą różnic czasowych 607

Q-uczenie 609

Strategie poszukiwania 610

Przybliżający algorytm Q-uczenia i Q-uczenie głębokie 611

Implementacja modelu Q-uczenia głębokiego 612

Odmiany Q-uczenia głębokiego 616

Ustalone Q-wartości docelowe 616

Podwójna sieć DQN 617

Odtwarzanie priorytetowych doświadczeń 618

Walcząca sieć DQN 618

Biblioteka TF-Agents 619

Instalacja biblioteki TF-Agents 620

Środowiska TF-Agents 620

Specyfikacja środowiska 621

Funkcje opakowujące środowisko i wstępne przetwarzanie środowiska Atari 622

Architektura ucząca 625

Tworzenie Q-sieci głębokiej 627

Tworzenie agenta DQN 629

Tworzenie bufora odtwarzania i związanego z nim obserwatora 630

Tworzenie wskaźników procesu uczenia 631

Tworzenie sterownika 632

Tworzenie zestawu danych 633

Tworzenie pętli uczenia 636

Przegląd popularnych algorytmów RN 637

Ćwiczenia 639

19. Wielkoskalowe uczenie i wdrażanie modeli TensorFlow 641

Eksploatacja modelu TensorFlow 642

Korzystanie z systemu TensorFlow Serving 642

Tworzenie usługi predykcyjnej na platformie GCP AI 650

Korzystanie z usługi prognozowania 655

Wdrażanie modelu na urządzeniu mobilnym lub wbudowanym 658

Przyspieszanie obliczeń za pomocą procesorów graficznych 661

Zakup własnej karty graficznej 662

Korzystanie z maszyny wirtualnej wyposażonej w procesor graficzny 664

Colaboratory 665

Zarządzanie pamięcią operacyjną karty graficznej 666

Umieszczanie operacji i zmiennych na urządzeniach 669

Przetwarzanie równoległe na wielu urządzeniach 671

Uczenie modeli za pomocą wielu urządzeń 673

Zrównoleglanie modelu 673

Zrównoleglanie danych 675

Uczenie wielkoskalowe za pomocą interfejsu strategii rozpraszania 680

Uczenie modelu za pomocą klastra TensorFlow 681

Realizowanie dużych grup zadań uczenia za pomocą usługi Google Cloud AI Platform 684

Penetracyjne strojenie hiperparametrów w usłudze AI Platform 686

Ćwiczenia 688

Dziękuję! 688

A. Rozwiązania ćwiczeń 691

B. Lista kontrolna projektu uczenia maszynowego 725

C. Problem dualny w maszynach wektorów nośnych 731

D. Różniczkowanie automatyczne 735

E. Inne popularne architektury sieci neuronowych 743

F. Specjalne struktury danych 751

G. Grafy TensorFlow 757