Sadržaj

Predgovor ..... **xv**

Deo I Osnove mašinskog učenja

1. Pejzaž mašinskog učenja **3**

Šta je mašinsko učenje? 4
Zašto koristiti mašinsko učenje? 4
Primeri primene 7
Vrste sistema mašinskog učenja 9
Nadgledano/nenadgledano učenje 9
Paketno i onlajn učenje 16
Učenje na primerima nasuprot učenju na modelu 19
Glavni izazovi mašinskog učenja 25
Nedovoljna količina podataka za obuku 26
Nereprezentativni podaci za obuku 27
Podaci lošeg kvaliteta 28
Nebitna svojstva 29
Preterano uklapanje podataka za obuku 29
Nedovoljno uklapanje podataka za obuku 31
Korak unazad 32
Testiranje i procena 32
Podešavanje hiperparametara i izbor modela 33
Neuklapanje podataka 34
Vežbe 35

2. Projekat mašinskog učenja od početka do kraja **37**

Rad sa stvarnim podacima 37
Sagledavanje celine 39
Definisanje problema 39
Izbor mere performanse 41
Provera pretpostavki 44
Pribavljanje podataka 44
Izrada radnog prostora 44
Preuzimanje podataka 47

Pogled na strukturu podataka 49
Izrada probnog skupa 52
Otkrivanje i vizuelizovanje podataka radi sticanja uvida 57
Vizuelizovanje geografskih podataka 57
Traženje korelacija 59
Eksperimentisanje sa kombinacijama atributa 62
Priprema podataka za algoritme mašinskog učenja 63
Prečišćavanje podataka 64
Rad sa tekstualnim i kategorijskim atributima 66
Namenski transformeri 68
Skaliranje svojstava 69
Transformacijski cevovodi 70
Izbor i obuka modela 72
Obuka i procena na trening skupu 72
Bolja procena unakrsnom proverom 73
Fino podešavanje modela 76
Rešetkasto pretraživanje 76
Nasumično pretraživanje 78
Metode ansambla 78
Analiza najboljih modela i njihovih grešaka 78
Procenite sistem na probnom skupu 79
Puštanje u pogon, praćenje i održavanje sistema 80
Pokušajte sami! 83
Vežbe 84

3. Klasifikacija 85
MNIST 85
Obučavanje binarnog klasifikatora 87
Mere performanse 88
Merenje tačnosti unakrsnom proverom 88
Matrica zabuna 90
Preciznost i prepoznavanje 92
Kompromis između preciznosti i prepoznavanja 93
ROC kriva 96
Višeklasna klasifikacija 99
Analize grešaka 102
Klasifikacija sa više oznaka 105
Klasifikacija sa više izlaza 106
Vežbe 107

4. Obučavanje modela 109
Linearna regresija 110
Normalna jednačina 111
Složenost proračuna 114

Gradijentni spust 115
Paketni gradijentni spust 118
Stohastički gradijentni spust 121
Mini-paketni gradijentni spust 124
Polinomska regresija 125
Krive učenja 127
Regularizovani linearni modeli 130
Grebenasta regresija 131
Lasso regresija 133
Elastična mreža 135
Rano zaustavljanje 136
Logistička regresija 138
Procenjivanje verovatnoća 138
Obuka i funkcija gubitka 139
Granice odlučivanja 140
Softmax regresija 142
Vežbe 146

5. Mašine sa vektorima podrške ... 147

Linearna SVM klasifikacija 147
Klasifikacija sa propustljivom marginom 148
Nelinearna SVM klasifikacija 150
Polinomsko jezgro 151
Svojstva sličnosti 152
Gausovo RBF jezgro 153
Složenost izračunavanja 155
SVM regresija 155
Ispod haube 157
Funkcija odlučivanja i predviđanja 157
Cilj obuke 158
Kvadratno programiranje 160
Dualni problem 161
Kernerlizovani SVM 161
Onlajn SVM 164
Vežbe 165

6. Stabla odluka ... 167

Obuka i vizuelizovanje stabla odluke 167
Predviđanja 168
Procenjivanje verovatnoće klase 170
Algoritam obuke CART 171
Složenost izračunavanja 172
Gini nečistoća ili entropija? 172
Hiperparametri regularizacije 173

Regresija 174
Nestabilnost 176
Vežbe 177

7. Učenje u ansamblu i nasumične šume 179
Klasifikatori glasanja 179
Samoagregacija i umetanje 182
Samoagregacija i umetanje u Scikit-Learnu 184
Procena oob 185
Nasumični isečci i nasumični podprostori 186
Nasumične šume 186
Ekstra stabla 187
Važnost svojstava 188
Pojačavanje 189
AdaBoost 189
Gradijentno pojačavanje 193
Nagomilavanje 197
Vežbe 200

8. Redukcija dimenzionalnosti 203
Prokletstvo dimenzionalnosti 204
Glavni načini redukcije dimenzionalnosti 205
Projekcija 205
Učenje mnogostrukosti 208
PCA 209
Očuvanje varijanse 209
Glavne komponente 210
Redukovanje na d dimenzija projektovanjem 211
Upotreba Scikit-Learna 212
Stopa objašnjene varijanse 212
Izbor tačnog broja dimenzija 212
PCA za komprimovanje 213
Nasumični PCA 214
Inkrementalni PCA 215
Kernel PCA 215
Izbor jezgra i podešavanje hiperparametara 216
LLE 219
Ostale tehnike redukcije dimenzionalnosti 220
Vežbe 222

9. Tehnike nenadgledanog učenja 223
Raspodela u klastere 224
K-Means 226
Ograničenja K-Meansa 236

Segmentacija slike klasterovanjem 236
Predobrada klasterovanjem 238
Polunadgledano učenje klasterovanjem 240
DBSCAN 242
Ostali algoritmi klasterovanja 245
Gausove mešavine 246
Otkrivanje anomalija pomoću Gausovih mešavina 252
Biranje broja klastera 253
Bajesovi modeli Gausovih mešavina 256
Ostali algoritmi za otkrivanje anomalija i noviteta 260
Vežbe 261

Deo II Neuronske mreže i duboko učenje

10. Uvod u veštačke neuronske mreže sa Kerasom ... 265
Od bioloških do veštačkih neurona 266
Biološki neuroni 267
Logička izračunavanja pomoću neurona 268
Perceptron 269
Višeslojni perceptron i povratna propagacija 273
MLP regresija 277
MLP klasifikacija 278
Implementiranje MLP-ova pomoću Kerasa 280
Instaliranje TensorFlowa 2 281
Izrada klasifikatora slika pomoću API-ja Sequential 282
Izrada regresionog MLP-a pomoću API-ja Sequential 291
Izrada složenih modela pomoću API-ja Functional 292
Upotreba API-ja Subclassing za izradu dinamičkih modela 296
Upisivanje i učitavanje modela 298
Upotreba callbacks 298
Upotreba TensorBoarda za vizuelizaciju 300
Fino podešavanje hiperparametara neuronskih mreža 303
Broj skrivenih slojeva 306
Broj neurona po skrivenom sloju 307
Brzina učenja, veličina paketa i ostali hiperparametri 308
Vežbe 310

11. Obuka dubokih neuronskih mreža ... 313
Problemi nestajućih/eksplodirajućih gradijenata 314
Glorotova i Heova inicijalizacija 315
Nezasićene aktivirajuće funkcije 316
Normalizacija paketa 320
Odsecanje gradijenta 326

Višekratna upotreba predobučenih slojeva 327
Prenosno učenje u Kerasu 328
Nenadgledana predobuka 330
Predobuka na pomoćnom zadatku 331
Brži optimizatori 332
Optimizacija momenta 332
Nesterovljev ubrzani gradijent 334
AdaGrad 335
RMSProp 336
Adam i Nadam optimizacija 337
Raspored brzine učenja 340
Izbegavanje prekomernog uklapanja pomoću regularizacije 344
ℓ_1 i ℓ_2 regularizacija 344
Izostavljanje 345
Monte Carlo (MC) izostavljanje 348
Max-Norm regularizacija 350
Rezime i praktične smernice 351
Vežbe 352

12. Namenski modeli i obuka pomoću TensorFlowa..... 355
Kratak pregled TensorFlowa 355
Upotreba TensorFlowa kao NumPya 358
Tenzori i operacije 358
Tenzori i NumPy 360
Konverzije tipova 361
Promenljive 361
Ostale strukture podataka 362
Prilagođavanje modela i algoritama obuke 363
Namenske funkcije gubitka 363
Upisivanje i učitavanje modela koji sadrže namenske komponente 364
Namenske aktivirajuće funkcije, inicijalizatori, regularizatori i ograničenja 366
Namenski merni pokazatelji 367
Namenski slojevi 370
Prilagođeni modeli 373
Gubici i merni pokazatelji zasnovani na unutrašnjosti modela 375
Izračunavanje gradijenata pomoću autodiffa 376
Namenske petlje za obuku 380
TensorFlow funkcije i grafovi 383
AutoGraph i iscrtavanje 384
Pravila TF funkcija 386
Vežbe 387

13. Učitavanje i predobrada podataka u TensorFlowu 389
API Data 390
Ulančavanje transformacija 390
Mešanje podataka 392
Predobrada podataka 395
Povezivanje svega zajedno 396
Predučitavanje 396
Upotreba skupa podataka u tf.kerasu 398
Format TFRecord 399
Komprimovane TFRecord datoteke 400
Kratak uvod u bafere protokola 400
TensorFlow protobaferi 402
Učitavanje i raščlanjivanje protobafera Example 403
Rad sa listama lista pomoću protobafera SequenceExample 404
Predobrada ulaznih osobina 405
Kodiranje kategorijskih osobina u obliku vektora jedan-vruće 406
Kodiranje kategorijskih osobina umetcima 408
Kerasovi slojevi predobrade 412
TF Transform 413
Projekat TensorFlow Datasets (TFDS) 415
Vežbe 416

14. Doboki kompjuterski vid upotrebom konvolucionih neuronskih mreža 419
Arhitektura vizuelnog korteksa 420
Konvolucioni slojevi 421
Filtri 423
Slaganje više mapa osobina 424
Implementacija u TensorFlowu 426
Memorijski zahtevi 428
Redukcioni slojevi 429
Implementacija u TensorFlowu 431
Arhitekture CNN 433
LeNet-5 435
AlexNet 436
GoogLeNet 438
VGGNet 442
ResNet 442
Xception 445
SENet 447
Implementiranje CNN ResNet-34 pomoću Kerasa 449
Upotreba predobučenih modela iz Kerasa 450
Predobučeni modeli za prenosno učenje 452
Klasifikacija i lokalizacija 454

Prepoznavanje objekata 456
Potpuno konvolucione mreže 457
YOLO (You Only Look Once) 459
Semantičko segmentovanje 462
Vežbe 466

15. Obrada sekvenci pomoću RNN i CNN 467
Rekurentni neuroni i slojevi 468
Memorijske ćelije 470
Ulazne i izlazne sekvence 471
Obuka RNN 472
Predviđanje vremenskih serija 473
Osnovni merni pokazatelji 475
Implementiranje jednostavne RNN 475
Duboke RNN 476
Predviđanje nekoliko vremenskih koraka unapred 478
Obrada dugačkih sekvenci 481
Borba sa problemom nestabilnih gradijenata 481
Rešavanje problema kratkoročne memorije 484
Vežbe 492

16. Obrada govornog jezika pomoću RNN i pažnje 493
Generisanje šekspirovskog teksta pomoću znakovnog RNN 494
Stvaranje skupa podataka za obuku 494
Kako podeliti sekvencijalni skup podataka 495
Isecanje sekvencijalnog skupa podataka u više prozora 496
Izrada i obučavanje modela Char-RNN 498
Upotreba modela Char-RNN 499
Generisanje izmišljenog šekspirovskog teksta 499
RNN sa stanjem 500
Analiza mišljenja 502
Maskiranje 506
Višekratna upotreba predobučenih umetaka 507
Mreža enkoder–dekoder za neuronsko mašinsko prevođenje 509
Dvosmerne RNN 512
Pretraživanje snopom 513
Mehanizmi pažnje 515
Vizuelna pažnja 518
Pažnja je sve što vam treba: Arhitektura Transformer 519
Nedavne inovacije u modelima jezika 527
Vežbe 529

17. Reprezentativno učenje i generativno učenje pomoću automatskih kodera i GAN-ova 531
Efiksno predstavljanje podataka 532
Izvršavanje PCA pomoću nepotpunog linearnog autoenkodera 534
Naslagani autoenkoderi 535
Implementiranje naslaganog autoenkodera u Kerasu 536
Prikazivanje rekonstrukcija 537
Vizuelizovanje skupa Fashion MNIST 538
Nenadgledana predobuka upotrebom naslaganih autoenkodera 539
Povezivanje težina 540
Obučavanje jednog po jednog autoenkodera 541
Konvolucioni autoenkoderi 543
Rekurentni autoenkoderi 543
Autoenkoderi za uklanjanje šuma 544
Proređeni autoenkoderi 546
Varijacioni autoenkoderi 549
Generisanje Fashion MNIST slika 553
Generativne suparničke mreže 554
Teškoće obučavanja GAN mreža 558
Duboke konvolucionione GAN mreže 559
Progresivni rast GAN mreža 563
StyleGAN mreže 565
Vežbe 568

18. Forsirano učenje 569
Učenje optimizacije nagrada 570
Nalaženje pravila 571
Uvod u OpenAI Gym 573
Pravila ponašanja neuronske mreže 577
Procenjivanje akcija: problem raspodele zasluga 578
Gradijenti pravila ponašanja 580
Markovljevi procesi odlučivanja 584
Učenje vremenske razlike 588
Q-učenje 589
Pravila istraživanja 591
Aproksimativno Q-učenje i duboko Q-učenje 591
Implemenatacija dubokog Q-učenja 592
Varijante dubokog Q-učenja 597
Fiksne ciljne Q-vrednosti 597
Dvostruki DQN 597
Prioriterni iskustveni odgovori 598
Duel DQN 599

Biblioteka TF-Agents 600
Instaliranje biblioteke TF-Agents 601
TF-Agents okruženja 601
Specifikacije okruženja 602
Omotači okruženja i predobrada Atari 603
Arhitektura obuke 606
Izrada duboke Q-mreže 607
Izrada DQN agenta 609
Izrada bafera odgovora i odgovarajućeg posmatrača 610
Izrada mernih pokazatelja obuke 612
Izrada drajvera za prikupljanje 613
Formiranje skupa podataka 614
Izrada petlje obuke 617
Pregled nekih popularnih algoritama forsiranog učenja 618
Vežbe 620

19. Obuka i uvođenje u primenu modela TensorFlow po meri 621
TensorFlow model u obliku usluge 622
Upotreba TensorFlow Servinga 622
Izrada servisa predviđanja na platformi GCP AI 631
Upotreba servisa predviđanja 634
Primena modela na mobilni ili ugrađen uređaj 637
Upotreba GPU za ubrzavanje izračunavanja 641
Nabavljanje GPU 642
Upotreba virtuelne mašine sa GPU 644
Saradnja 645
Upravljanje RAM-om GPU-a 646
Raspoređivanje operacija i promenljivih na uređaje 649
Paralelno izvršavanje na više uređaja 651
Obuka modela pomoću više uređaja 653
Paralelizam modela 653
Paralelizam podataka 655
Obuka po meri pomoću API-ja Distribution Strategies 660
Obuka modela na TensorFlow klasteru 661
Izvršavanje obimnih poslova obuke na Google Cloud AI platformi 664
Podešavanje hiperparametara crnom kutijom na AI platformi 666
Vežbe 667
Hvala! 668

Dodaci

A Rešenja vežbi 669

B Kontrolna lista projekta mašinskog učenja 703

C Problem dualnosti SVM 709

D Autodiff 713

E Ostale popularne arhitekture veštačkih neuronskih mreža 721

F Specijalne strukture podataka 731

G TensorFlow grafovi 737

Rečnik termina korišćenih u knjizi 745

Indeks 753

Predgovor

Cunami mašinskog učenja

2006. godine Geoffrey Hinton i dr, objavili su rad (*https://homl.info/136*)[1] u kojem pokazuju kako se obučava duboka neuronska mreža sposobna da rukom napisane cifre prepoznaje s vrhunskom preciznošću (> 98%). Ovu tehniku su nazvali „Duboko učenje". Duboka neuronska mreža je (veoma) pojednostavljen model naše moždane kore, sastavljen od slojeva veštačkih neurona. Obučavanje duboke neuronske mreže tada se uglavnom smatralo nemogućim[2] i krajem 1990-ih je većina istraživača napustila tu ideju. Ovaj rad je oživeo interesovanje naučne zajednice i ubrzo mnogi novi radovi su pokazali da je duboko učenje ne samo moguće, nego i da je u stanju da postigne zadivljujuće rezultate koje nijedna druga tehnika mašinskog učenja (MU) ne može da pruži (pomoću ogromne računarske snage i velikih količina podataka). Taj entuzijazam se ubrzo proširio i na mnoga druga područja mašinskog učenja.

Deceniju ili nešto kasnije, mašinsko učenje je pokorilo industriju: ono čini srce najvećeg dela čarolije današnjih visokotehnoloških proizvoda i rangira rezultate pretraživanja veba, omogućava da vaš pametni telefon prepoznaje govor, preporučuje video snimke i pobeđuje svetskog prvaka u igri Go. Čak i pre nego što postanete svesni toga, upravljaće i vašim automobilom.

Mašinsko učenje u vašim projektima

Sasvim je prirodno da budete uzbuđeni zbog mašinskog učenja i da želite da se pridružite zabavi!

Možda biste želeli da ugradite mozak robotu koji ste sami napravili? Učiniti da prepoznaje lica? Ili naučiti ga da hoda okolo?

[1] Geoffrey E. Hinton i dr. „A Fast Learning Algorithm for Deep Belief Nets", Neural Computation 18 (2006): 1527–1554.

[2] Uprkos činjenici da su duboke konvolucione neuronske mreže Yann LeCuna dobro funkcionisale za prepoznavanje slika još od devedesetih godina prošlog veka, mada one nisu bile opšte namene.

Ili možda vaša kompanija ima gomile podataka (podaci o aktivnostima korisnika, finansijski podaci, podaci o proizvodnji, podaci iz senzora na mašinama, statistike pozivanja kolcentra, izveštaji o zaposelnima itd.) i vrlo je verovatno da biste tu mogli da otkrijete neke skrivene dragulje kada biste samo znali gde da ih tražite. Pomoću mašinskog učenja možete postići sledeće, a i još više (*https://homl.info/usecases*):

- Raspodela kupaca u segmente i pronalaženje najbolje marketinške strategije za svaki segment.
- Preporučivanje proizvoda svakom klijentu na osnovu onoga što su kupili slični klijenti.
- Otkrivanje transakcija koje mogu biti lažne.
- Prognoza prihoda za narednu godinu.

Bez obzira na razlog, odlučili ste da naučite mašinsko učenje i implementirate ga u svoje projekte. Odlična ideja!

Cilj i pristup

Ova knjiga pretpostavlja da vam nije poznato praktično ništa o mašinskom učenju. Njen cilj je da vam pruži koncepte, alate i intuiciju koji su vam potrebni da biste implementirali programe koji su u stanju da *uče iz podataka*.

Razmotrićemo veliki broj tehnika, od najjednostavnijih i nauobičajenijih (kao što je linearna regresija) do nekih tehnika dubokog učenja koje redovno pobeđuju na takmičenjima.

Umesto da implementiramo vlastite jednostavne verzije svakog algoritma, koristićemo Pythonove postojeće radne okvire:

- Mada je biblioteka Scikit-Learn (*http://scikit-learn.org/*) vrlo jednostavna za upotrebu, efikasno implementira mnoge algoritme mašinskog učenja, tako da predstavlja sjajnu početnu tačku za mašinsko učenje. Tu biblioteku je osmislio David Cournapeau 2007. godine, a danas je održava tim istraživača na fransuskom Institutu za istraživanje računarskih nauka i automatizaciju (Inria).
- TensorFlow (*https://tensorflow.org/*) je složenija biblioteka za distribuirane numeričke proračune. Omogućava efikasno obučavanje i upotrebu veoma obimnih neuronskih mreža tako što izračunavanja može da rasporedi čak na stotine servera opremljenih GPU jedinicama (jedinice za grafičku obradu). TensorFlow (TF) je osmišljen u kompaniji Google i podržava mnoge njihove aplikacije za mašinsko učenje širokog opsega. Od novembra 2015 je na raspolaganju kao otvoren kôd, a septembra 2019. objavljena je verzija 2.0.
- Keras (*https://keras.io/*) je API visokog nivoa za duboko učenje koji čini vrlo jednostavnim obučavanje i upotrebu neuronskih mreža. Može raditi u kombinaciji s TensorFlowom, Theanom ili Microsoftovim Cogitive Toolkitom (ranije poznat kao CNTK). TensorFlow se isporučuje sa svojom sopstvenom implementacijom ovog API-ja, nazvan tf.keras, koja pruža podršku za neke napredne funkcije TensorFlowa (npr. mogućnost efikasnog učitavanja podataka).

Ova knjiga je osmišljena kao praktičan pristup, da biste razvili intuitivno razumevanje mašinskog učenja kroz konkretne primere i samo malo teorije. Iako ovu knjigu možete čitati bez upotrebe računara, preporučujem vam da eksperimentišete s primerima koda koji su dostupni na vebu u formatu Jupyter beležnica na *https://github.com/ageron/handson-ml2*.

Predznanja

Ova knjiga pretpostavlja da imate bar neko iskustvo sa programiranjem na jeziku Python i da poznajete glavne Pythonove naučne biblioteke - naročito NumPy (*http://numpy.org/*), pandas (*http://pandas.pydata.org/*) i Matplotlib (*http://matplotlib.org/*).

Osim toga, ako vas zanima šta je „ispod haube", treba vam solidno znanje iz srednjoškolske matematike (infinitezimalni račun, linearna algebra, verovatnoća i statistika).

Ako ne znate Python, možete krenuti od knjiga: Python: uvod u programiranje (*https://www.mikroknjiga.rs/store/prikaz.php?ref=978-86-7555-362-5*), Naučite programiranje (*https://www.mikroknjiga.rs/store/prikaz.php?ref=978-86-7555-430-1*) ili veb lokacije *http://learnpithon.org*. Zvanični tutorial na Python.org je takođe prilično dobar.

Ako nikada niste koristili Jupyter, poglavlje 2 će vas voditi kroz instalaciju i osnove: to je moćna alatka koju treba imati u svojoj kolekciji alatki.

Ako ne poznajete Pythonove naučne biblioteke, postoje Jupyter beležnice s uputstvima i primerima. Postoji i kratak matematički vodič za linearnu algebru.

Mapa puta

Ova knjiga je organizovana u dva dela. Prvi deo, *Osnove mašinskog učenja*, obrađuje sledeće teme:

- Šta je mašinsko učenje, koje probleme pokušava da reši i koje su glavne kategorije i osnovni koncepti tog sistema
- Koraci jednog tipičnog projekta mašinskog učenja
- Učenje uklapanjem modela u podatke
- Optimizacija funkcije gubitka
- Rukovanje, prečišćavanje i priprema podataka
- Biranje i prilagođavanje osobina podataka
- Odabir modela i podešavanje hiperparametara pomoću unakrsne provere
- Izazovi mašinskog učenja, naročito nedovoljno i prekomerno uklapanje modela (kompromis između odstupanja i varijanse)
- Najuobičajeniji algoritmi za učenje: linearna i polinomska regresija, logistička regresija, K-najbližih suseda, mašine se vektorima podrške, stablo odlučivanja, nasumične šume i metode ansambla

- Smanjivanje dimenzija podataka za obuku u borbi protiv „prokletstva dimenzionalnosti“
- Ostale nenadzirane tehnike učenja, među koima su grupisanje, procena gustine i otkrivanje anomalija

Deo II, *Neuronske mreže i duboko učenje*, razmatra sledeće teme:

- Šta su neuronske mreže i za šta su one dobre
- Izrada i obuka neuronskih mreža pomoću TensorFlowa i Kerasa
- Najvažnije arhitekture neuronskih mreža: feedforward neuronske mreže za tabelarne podatke, konvolucione mreže za računarski vid, rekurentne mreže i duge kratkoročne mreže (LSTM) za obradu sekvenci, koderi/dekoderi i transformeri za obradu govornog jezika, autoenkoderi i generativne suparničke mreže (GAN) za generativno učenje
- Tehnike za obuku dubokih neuronskih mreža
- Kako da napravite agent (npr. bot za igre) koji metodom pokušaja i grešaka može da nauči dobre strategije primenom forsiranog učenja
- Efikasno učitavanje i obrada velikih količina podataka
- Obuka i uvođenje u upotrebu TensorFlow modela prema potrebama

U prvom delu koristimo uglavnom biblioteku Scikit-Learn, dok u drugom delu koristimo TensorFlow i Keras.



Nemojte prebrzo skakati u duboku vodu: dok je duboko učenje bez sumnje jedno od najuzbudljivijih područja mašinskog učenja, prvo bi trebalo savladati osnove. Štaviše, većina problema se može prilično uspešno rešiti upotrebom jednostavnijih tehnika, kao što su nasumične šume i metoda ansambla (o kojima je reč u delu I). Duboko učenje je najprikladnije za složene probleme kao što su prepoznavanje slika, razumevanje govora ili obrada govornog jezika, pod uslovom da imate dovoljno podataka, računarske snage i strpljenja.

Izmene u drugom izdanju

Ovo drugo izdanje ima šest glavnih ciljeva:

1. Razmatranje dodatnih tema u vezi s MU: dodatne nenadgledane tehnike učenja (kao što su grupisanje, otkrivanje anomalija, procena gustine i modeli mešanja); više tehnika za obuku dubokih mreža (uključujući samonormalizujuće mreže); dodatne tehnike računarskog vida (uključujući Xception, SENet, prepoznavanje predmeta pomoću YOLO i semantička segmentacija pomoću R-CNN); obrada sekvenci pomoću konvolucionih neuronskih mreža (KNM, uključujući WaveNet); obrada govornog jezika pomoću rekurentnih neuronskih mreža (RNM), KNM i transformera i GAN mreže.

2. Razmatranje dodatnih biblioteka i API-ja (Keras, Data API, TF agenti za forsirano učenje) i obuka i uvođenje u upotrebu TF modela pomoću API-ja Distribution Strategies, TF-Serving servera i Google Cloud AI platforme. Takođe kratko predstavljanje TF Transforma, TFLitea, TF Addons/Sek2Seka i TensorFlow.js.
3. Diskusija o nekim najnovijim važnim rezultatima istraživanja u oblasti mašinskog učenja.
4. Prilagođavanje na TensorFlow 2 svih poglavlja gde se koristi TensorFlow i upotreba TensorFlowove implementacije API-ja Keras (tf.keras) gde god je to moguće.
5. Ažuriranje primera koda da biste koristili najnovije verzije bilioteka Scikit-Learn, NumPy, pandas, Matplotlib i drugih biblioteka.
6. Pojašnjavanje nekih odeljaka i ispravljanje nekih grešaka, zahvaljujući sjajnim povratnim informacijama od čitalaca.

Dodata su neka nova poglavlja, druga su prepravljena, a neka su poređana drugim redosledom. Videti na *https://homl.info/changes2* detaljniji opis onoga što je izmenjeno u drugom izdanju.

Ostali resursi

Na raspolaganju su brojni odlični resursi za učenje mašinskog učenja. Na primer, kurs Andrew Nga na Courseri (*https://homl.info/ngcourse*) je neverovatan, iako zahteva značajno ulaganje vremena (računajte više meseci).

Imate i mnogo zanimljivih veb lokacija o mašinskom učenju, među kojima je i izuzetan priručnik za korisnike biblioteke Scikit-Learn (*https://homl.info/skdoc*). Možda će vam se svideti i Dataquest (*https://www.dataquest.io/*) koji nudi veoma lepe interaktivne tutoriale, kao i ML blogovi poput onih koji su navedeni na sajtu Quora (*https://homl.info/1*). I konačno, veb lokacija Deep Learning (*http://deeplearning.net/*) ima dobru listu resursa.

Postoje i mnoge druge uvodne knjige o mašinskom učenju. Naročito:

- *Data Science from Scratch*, Joela Grusa (izdavač je O'Reilly) predstavlja osnove mašinskog učenja i implementira neke od glavnih algoritama u čistom Pythonu (od nule, kao što ime upućuje).
- *Machine Learning: An Algorithmic Perspective*, Stephena Marslanda (izdavač je Chapman & Hall) je sjajan uvod u mašinsko učenje, koji detaljno obrađuje širok opseg tema s primerima koda u Pythonu (takođe od početka, ali s bibliotekom NumPy).
- *Python Machine Learning*, Sebastiana Raschkea (izdavač je Packt Publishing) takođe je odličan uvod u mašinsko učenje i koristi Python biblioteke otvorenog koda (Pilearn 2 i Theano).
- *Deep Learning with Python*, Francoisa Cholleta (izdavač je Manning) vrlo je praktična knjiga koja na jasan i sažet način obrađuje veliki broj tema, kao što biste mogli i očekivati od autora odlične biblioteke Keras. Daje prednost primerima koda ispred matematičke teorije.

- *The Hundred-Page Machine Learning Book*, knjiga Andreja Burkova veoma je kratka i sadrži impresivan opseg tema, koje uvodi na pristupačan način, ali ne izbegavajući matematičke jednačine.
- *Learning from Data*, čiji su autori Yaser S. Abu-Mostafa, Malik Magdon-Ismail i Hsuan-Tien Lin (izdavač je AMLBook) više je teoretski pristup mašinskom učenju koji pruža duboke uvide, posebno o kompromisu varijansa/odstupanje (videti poglavlje 4).
- *Artificial Intelligence: A Modern Approach*, treće izdanje, čiji su autori Stuart Russel i Peter Norvig (izdavač je Pearson), sjajna je (i ogromna) knjiga koja obrađuje neverovatnu količinu tema, uključujući i mašinsko učenje.

I konačno, učlanjivanje na neku veb lokaciju za takmičenje u mašinskom učenju, kao što je Kaggle.com, omogućiće vam da vežbate svoje veštine na stvarnim problemima, uz pomoć i podršku nekih od najboljih profesionalaca u oblasti MU.

Konvencije koje se koriste u ovoj knjizi

U ovoj knjizi se koriste sledeće tipografske konvencije:

Kurzivna slova

Označavaju nove pojmove, URL adrese, adrese e-pošte, imena datoteka i nastavke imena datoteka.

`Slova fiksne širine`

Koriste se za listinge programa, kao i u unutar teksta za označavanje elemenata programa kao što su imena promenljivih ili funkcija, baza podataka, tipova podataka, promenljivih okruženja, naredbe i ključne reči.

`Podebljana slova fiksne širine`

Prikazuju komande ili drugi tekst koji korisnik treba da upiše doslovno.

`Kurzivna slova fiksne širine`

Prikazuju tekst koji korisnik treba da zameni svojim vrednostima ili vrednostima koje određuje kontekst.



Ovaj element označava savet ili predlog.



Ovaj element označava napomenu.



Ovaj element označava upozorenje ili oprez.

Primeri koda

Postoji niz Jupyterovih beležnica prepunih dopunskog materijala, kao što su primeri koda i vežbe, koje možete preuzeti na *https://github.com/ageron/handson-ml2*.

Neki primeri koda u knjizi izostavljaju delove koji se ponavljaju ili detalje koji su očigledni ili nisu ni kakvoj vezi s mašinskim učenjem. Tako se zadržava usredsređenost na važne delove koda i štedi prostor da bi se obuhvatio veći broj tema. Ako želite kompletne primere koda, svi su na raspolaganju u obliku Jupyterovih beležnica.

Imajte na umu da kada primeri koda prikazuju rezultate, ti primeri koda prikazani su s Pythonovim odzivnicima (`>>>` i `...`), kao u Python ljusci, da biste jasno razlikovali kôd od rezultata. Na primer, ovaj kôd definiše funkciju `square()`, zatim izračunava i prikazuje vrednost kvadrata od 3:

```
>>> def square(x):
...     return x ** 2
...
>>> result = square(3)
>>> result
9
```

Kada kôd ne prikazuje nikakav rezultat, odzivnici se ne koriste. Međutim, rezultat se ponekad može prikazati i u obliku komentara, kao što je ovaj:

```
def square(x):
    return k ** 2

result = square(3)  # rezultat je 9
```

Korišćenje primera koda

Namena ove knjige je da vam pomogne da obavite svoj posao. Ako je u knjizi naveden primer, uglavnom ga možete koristiti u svojim programima i dokumentaciji. Ne morate nam tražiti dozvolu ako ne reprodukujete značajan deo koda. Na primer, za pisanje programa koji koristi više blokova koda iz ove knjige nije potrebno odobrenje. Umetanje značajnog dela primera koda iz ove knjige u dokumentaciju vašeg proizvoda zahteva prethodnu dozvolu.

Zahvalnost

Ni u svojim najluđim snovima nisam zamišljao da će prvo izdanje ove knjige privući tako veliku publiku. Dobio sam toliko poruka od čitalaca, u kojima su mnogi postavljali pitanja, neki ljubazno isticali greške, a većina mi je poslala ohrabrujuće reči. Ne mogu izraziti koliko sam zahvalan svima njima na ogromnoj podršci. Hvala svima! Molim vas da ako otkrijete greške u primerima koda (ili samo da biste postavili pitanja) ne ustručavate se da ih prijavite na GitHub (*https://homl.info/issues2*) ili pošaljete ispravke tih grešaka (*https://homl.info/errata2*). Neki čitaoci su takođe podelili kako im je ova knjiga pomogla da pronađu svoj prvi posao ili kako im je pomogla da reše konkretan problem na kojem su radili. Smatram da su takve povratne informacije neverovatno motivirajuće. Ako smatrate da vam je ova knjiga korisna, voleo bih kada biste svoju priču podelili sa mnom, privatno (npr. putem LinkedIna na *https://www.linkedin.com/in/aurelien-geron/*) ili javno (npr. u obliku tvita ili recenzije na Amazonu).

Takođe sam izuzetno zahvalan svim divnim ljudima koji su izdvojili vreme iz svojih užurbanih života da bi moju knjigu pregledali s tolikom pažnjom. Posebno bih se zahvalio Francois Cholletu na pregledu svih poglavlja o Kerasu i TensorFlowu i pružanju nekoliko sjajnih povratnih informacija. Budući da je Keras jedan od glavnih dodataka ovom drugom izdanju, recenzija od samog autora Kerasa za mene ima neprocenjivu vrednost. Toplo preporučujem Francoisovu knjigu *Deep Learning with Python* (*https://homl.info/cholletbook*) (izdavač je Manning): ona ima sažetost, jasnoću i dubinu same biblioteke Keras. Posebno sam zahvalan Ankur Patelu, koji je pregledao svako poglavlje ovog drugog izdanja i pružio mi odlične povratne informacije, posebno o poglavlju 9, koje obrađuje nenadgledane tehnike učenja. Mogao je napisati celu knjigu na tu temu… a, čekajte, već jeste! Pogledajte *Hands-On Unsupervised Learning Using Python: How to Build Applied Machine Learning Solutions from Unlabeled Data* (*https://homl.info/patel*), izdavač je O'Reilly. Veliko hvala i Olzhasu Akpambetovu, koji je pregledao sva poglavlja u drugom delu knjige, testirao veći deo koda i ponudio mnogo sjajnih predloga. Zahvalan sam Marku Daoustu, Jon Krohnu, Dominiku Monnu i Joshu Pattersonu na recenziji drugog dela ove knjige i uključivanju njihove stručnosti. Nisu ostavili nijedan kamen neprevrnut i pružili su neizmerno korisne povratne informacije.

Dok sam pisao ovo drugo izdanje, imao sam sreću da dobijem veliku pomoć od članova TensorFlow tima, naročito Martina Wickea, koji je neumorno odgovarao na desetine mojih pitanja a ostatala slao pravim ljudima, među kojima su Karmel Allison, Paige Bailey, Eugene Brevdo, William Chargin, Daniel „Wolff" Dobson, Nick Felt, Bruce Fontaine, Goldie Gadde, Sandeep Gupta, Priya Gupta, Kevin Haas, Konstantinos Katsiapis, Viacheslav Kovalevski, Allen Lavoie, Clemens Mewald, Dan Moldovan, Sean Morgan, Tom O'Malley, Alexandre Passos, Andre Susano Pinto, Anthony Platanios, Oscar Ramirez, Anna Revinskaya, Saurabh Saxena, Ryan Sepassi, Jiri Simsa, Xiaodan Song, Christina Sorokin, Dustin Tran, Todd Wang, Pete Warden (koji je recenzirao i prvo izdanje) Edd Wilder-James i Yuefeng Zhou, koji su svi bili od izuzetne pomoći. Veliko hvala svima vama i svim ostalim članovima TensorFlow tima, ne samo na vašoj pomoći, već i što ste napravili tako sjajnu biblioteku! Posebna zahvalnost Ireni Giannoumis i Robertu Croweu iz TFX tima za detaljni pregled poglavlja 13 i 19.

Veliko hvala i fantastičnom osoblju O'Reillya, posebno Nicole Taché, koja mi je davala vredne povratne informacije i uvek bila vesela, ohrabrujuća i korisna. Nisam mogao da sanjam o boljem uredniku. Veliko hvala i Michele Cronin, koja je bila od velike pomoći (i strpljenja) od samog početka ovog drugog izdanja, i Kristen Brown, urednici drugog izdanja, koja je prošla kroz sve korake (koordinisala je i popravke i ispravke za svaki reprint prvog izdanja). Hvala i Rachel Mongan i Amandi Kersei na njihovim temeljnim ispravkama (prvog, odnosno drugog izdanja), kao i Johnnyju O'Tooleu koji je bio veza sa Amazonom i odgovario na mnoga moja pitanja. Zahvaljujem se Marie Beaugureau, Benu Loricu, Mikeu Loukidesu i Laurelu Rumi što su verovali u ovaj projekat i pomogli mi da definišem njegov opseg. Hvala Mattu Hackeru i celom Atlas timu koji su odgovorili na sva moja tehnička pitanja u vezi sa formatiranjem u AsciiDocu i LaTeX-u, a hvala i Nick Adamsou, Rebecci Demarest, Rachel Head, Judith McConville, Helen Monroe, Karen Montgomeri, Rachel Roumeliotis i svima drugima u O'Reillyu koji su doprineli ovoj knjizi.

Želeo bih da se zahvalim i mojim ranijim kolegama iz kompanije Google, posebno YouTubeovom timu za klasifikaciju video snimaka, što su mi pokazali toliko mnogo toga u vezi s mašinskim učenjem. Bez njih ne bih mogao ni da započnem prvo izdanje ove knjige. Posebno se zahvaljujem mojim ličnim „guruima" za mašinsko učenje: Clementu Courbetu, Julienu Dubois, Mathiasu Kendeu, Danielu Kitachewskom, Jamesu Packu, Alexanderu Paku, Anoshu Raju, Vitoru Sessaku, Wiktoru Tomczaku, Ingrid von Glehn i Richu Washingtonu. Hvala i svim ostalim s kojima sam radio u YouTubeu, kao i izuzetnim Googleovim istraživačkim timovima u Mountain Viewu. Zahvaljujem se i Martinu Andrewsu, Samu Witteveenu i Jasonu Zamanu na dobrodošlici u njihovu grupu Google Developer Experts u Singapouru, uz ljubaznu podršku Soonsona Kwona, kao i za sve odlične diskusije koje smo imali u vezi s mašinskim učenjem i TensorFlowom. Svako u Singapuru koga zanima mašinsko učenje treba da učestvuje na sastancima njihove grupe Deep Learning Singapore (*https://homl.info/meetupsg*). Jason zaslužuje posebnu zahvalnost što mi je stavio na raspolaganje deo svog znanja o TFLiteu u poglavlju 19!

Neću nikad zaboraviti sve divne ljude koji su recenzirali prvo izdanje ove knjige, među kojima su David Andrzejewski, Lukas Biewald, Justin Francis, Vincent Guilbeau, Eddy Hung, Karim Matrah, Gregoire Mesnil, Salim Semaoune, Iain Smears, Michel Tessier, Ingrid von Glehn, Pete Warden i razume se, moj dragi brat Sylvain. Posebno se zahvaljujem Haesunu Parku, koji mi je pružio mnoge odlične povratne informacije i otkrio nekoliko grešaka dok je prevodio na korejski jezik prvo izdanje ove knjige. Preveo je na korejski i Jupyterove beležnice, da ni ne pominjem dokumentaciju TensorFlowa. Ja ne govorim korejski, ali sudeći po kvalitetu komunikacije s njim, mora da su svi njegovi prevodi zaista odlični! Haesun je ljubazno doprineo i nekim rešenjima u ovom drugom izdanju.

Poslednje ali ne i najmanje važno, beskrajno sam zahvalan mojoj voljenoj supruzi, Emmanuellei i našoj divnoj deci, Alexandreu, Remiju i Gabriellei, za svu podršku tokom rada na ovoj knjizi. Zahvalan sam im i na njihovoj neutoljivoj radoznalosti: objašnjavanje nekih među najtežim konceptima u ovoj knjizi mojoj ženi i deci pomoglo mi je da raščistim misli i direktno poboljšalo mnoge delove knjige. A i stalno su mi donosili kafu i kolačiće! Šta više može čovek da sanja?