

Opinie o książce *Generatywne głębokie uczenie*

Generatywne głębokie uczenie się jest dostępnym wprowadzeniem do zestawu narzędzi głębokiego uczenia się, związanych z generatywnym modelowaniem. Jeśli jesteście kreatywnymi praktykami kochającymi majstrować przy kodzie i chcecie zastosować głębokie uczenie się w waszej pracy, ta książka jest dla Was.

– *David Ha, szef strategii w Stability AI*

Doskonała książka, która zagłębia się prosto we wszystkie najważniejsze techniki leżące u podstaw współczesnego głębokiego uczenia się. Znajdziecie tu intuicyjne wyjaśnienia i mądre analogie – poparte dydaktycznymi, bardzo czytelnymi przykładami kodu. Pobudzająca analiza jednej z najbardziej fascynujących dziedzin w AI!

– *François Chollet, twórca Keras*

Wyjaśnienie przez Davida Fostera złożonych pojęć jest jasne i zwięzłe, wzbogacone intuicyjnymi wizualizacjami, przykładami kodu i ćwiczeniami. Doskonałe źródło dla studentów i praktyków!

– *Suzana Ilić, główny menedżer programowy Responsible AI w Microsoft Azure OpenAI*

Generatywna AI to kolejny rewolucyjny krok technologii AI, który będzie miał ogromny wpływ na cały świat. Książka daje wspaniałe wprowadzenie do tej dziedziny i jej niezwykły potencjał oraz ryzyko z nią związane.

– *Connor Leahy, prezes firmy Conjecture i współzałożyciel EleutherAI*

Przewidywanie świata oznacza jego zrozumienie – we wszystkich jego odmianach. W tym sensie generatywna AI rozwiązuje najbardziej podstawowe zagadnienia inteligencji.

– *Jonas Andrulis, założyciel i prezes Aleph Alpha*

Generatywna AI zmienia niezliczone branże i daje nową generację kreatywnych narzędzi. Ta książka stanowi doskonały sposób na zapoznanie się z generatywnym modelowaniem i samodzielne rozpoczęcie tworzenia z wykorzystaniem tej rewolucyjnej techniki.

– *Ed Newton-Rex, twórca w Stability AI i kompozytor*

David nauczył mnie wszystkiego, co wiem o uczeniu się maszyn i ma talent do tłumaczenia pojęć leżących u jego podstaw. *Generatywne głębokie uczenie się* jest moim wprowadzającym narzędziem do generatywnej AI i stoi na półce obok mojego biurka wraz z niewielkim zbiorem ulubionych książek technicznych.

– *Zack Thoutt, CPO w AutoSalesVelocity*

Generatywna AI ma szansę wyrzucić ogromny wpływ na społeczeństwo. Książka ta to wprowadzenie do tej dziedziny, dostępne bez skupiania się na szczegółach technicznych.

– *Raza Habib, współzałożyciel Humanloop*

Gdy ktoś mnie pyta, jak wejść w tematykę generatywnej AI, zawsze polecam książkę Davida. Jej drugie wydanie robi wrażenie, gdyż obejmuje najsilniejsze modele, jak modele dyfuzyjne i transformery. Pozycja niezbędna dla każdego, kto jest zainteresowany kreatywnością obliczeniową!

– *Dr. Tristan Behrens, ekspert i muzyk AI w firmie Residence w KI Salon Heilbronn*

Jest to mój numer jeden praktycznej literatury wypakowany wiedzą techniczną, gdy mam pomysły na temat generatywnej AI. Książka ta powinna być na półce każdego analityka danych.

– *Martin Musiol, twórca generativeAI.net*

Książka w wyśmienitych szczegółach obejmuje pełną taksonomię modeli generatywnych. Jedną z najlepszych rzeczy, których dowiedziałem się o tej książce, to fakt, że obejmuje ważną teorię leżącą u podstaw modeli, a także utrwała zrozumienie czytelnika za pomocą przykładów praktycznych. Muszę zwrócić uwagę, że rozdział na temat GAN jest jednym z najlepszych wyjaśnień, jakie czytałem i daje intuicyjne środki do dostrajania swoich modeli. Książka obejmuje szeroki zakres różnorodnych generatywnych AI, w tym teksty, muzykę i obrazy. Jest to idealne źródło wiedzy dla każdego, kto rozpoczyna działanie z generatywną AI.

– *Aishwarya Srinivasan, analityk danych, Google Cloud*

Wydanie 2

Generatywne głębokie uczenie

*Uczenie maszyn, jak malować,
pisać, komponować i grać*

David Foster
Słowo wstępne Karl Friston

przekład: Mariusz Rogulski, Witold Sikorski

APN Promise
Warszawa 2024

O'REILLY

Generatywne głębokie uczenie, wydanie 2

Authorized translation of English edition of
Generative Deep Learning, 2nd edition, by David Foster
ISBN 978-1-098-13418-1

Copyright © 2023 Applied Data Science Partners Ltd.
All rights reserved.

This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls of all rights to publish and sell the same.

Copyright wydania polskiego © 2024 APN Promise SA
APN PROMISE SA, ul. Domaniewska 44a, 02-672 Warszawa
tel. +48 22 35 51 600
e-mail: wydawnictwo@promise.pl

Wszystkie prawa zastrzeżone. Żadna część niniejszej książki nie może być powielana ani rozpowszechniana w jakiegokolwiek formie i w jakikolwiek sposób (elektroniczny, mechaniczny), włącznie z fotokopiuowaniem, nagrywaniem na taśmy lub przy użyciu innych systemów bez pisemnej zgody wydawcy.

Logo O'Reilly jest zarejestrowanym znakiem towarowym O'Reilly Media, Inc. Ilustracja z okładki i powiązane elementy są znakami towarowymi O'Reilly Media, Inc.

Wszystkie inne nazwy handlowe i towarowe występujące w niniejszej publikacji mogą być znakami towarowymi zastrzeżonymi lub nazwami zastrzeżonymi odpowiednich firm odnośnych właścicieli.

Przykłady firm, produktów, osób i wydarzeń opisane w niniejszej książce są fikcyjne i nie odnoszą się do żadnych konkretnych firm, produktów, osób i wydarzeń. Ewentualne podobieństwo do jakiegokolwiek rzeczywistej firmy, organizacji, produktu, nazwy domeny, adresu poczty elektronicznej, logo, osoby, miejsca lub zdarzenia jest przypadkowe i niezamierzone.

APN PROMISE SA dołożyła wszelkich starań, aby zapewnić najwyższą jakość tej publikacji. Jednakże nikomu nie udziela się rękojmi ani gwarancji.

APN PROMISE SA nie jest w żadnym wypadku odpowiedzialna za jakiegokolwiek szkody będące następstwem korzystania z informacji zawartych w niniejszej publikacji, nawet jeśli APN PROMISE została powiadomiona o możliwości wystąpienia szkód.

ISBN: 978-83-7541-548-3 (druk), 978-83-7541-549-0 (ebook)

Projekt okładki: Karen Montgomery
Ilustracje: Kate Dullea

Przekład: Mariusz Rogulski i Witold Sikorski
Redakcja: Marek Włodarz
Korekta: Ewa Swędrowska
Skład i łamanie: MAWart Marek Włodarz

Dla Aliny, najukochańszego wektora szumu spośród wszystkich.

Spis treści

Przedmowa	xv
Wstęp	xvii

Część I Wprowadzenie do generatywnego głębokiego uczenia

1. Modelowanie generatywne.....	3
Czym jest modelowanie generatywne?	4
Modelowanie generatywne vs. dyskryminacyjne	5
Powstanie modelowania generatywnego	6
Modelowanie generatywne i sztuczna inteligencja	8
Nasz pierwszy model generatywny	9
Witaj, świecie!	9
Ramy modelowania generatywnego	10
Uczenie reprezentacji	12
Podstawowa teoria prawdopodobieństwa	15
Taksonomia modelu generatywnego	18
Baza kodu dla generatywnego głębokiego uczenia	20
Klonowanie repozytorium	20
Użycie Dockera	20
Uruchamianie na GPU	20
Podsumowanie	21
2. Uczenie głębokie	23
Dane dla uczenia głębokiego	24
Głębokie sieci neuronowe	25
Czym jest sieć neuronowa?	25
Uczenie się wysokopoziomowych cech	26
TensorFlow i Keras	27
Perceptron wielowarstwowy (MLP)	28
Przygotowanie danych	28
Budowanie modelu	30
Kompilowanie modelu	35
Szkolenie modelu	36

Ocena modelu	38
Splotowa siec neuronowa (CNN).....	40
Warstwy splotowe	40
Wsadowa normalizacja.....	45
Odrzucanie	48
Budowanie CNN	50
Szkolenie oraz ocena CNN.....	52
Podsumowanie.....	53

Część II Metody

3. Autokodery wariacyjne.....	57
Wprowadzenie	58
Autokodery	59
Zbiór danych Fashion-MNIST.....	60
Architektura autokodera.....	61
Koder.....	62
Dekoder.....	63
Połączenie kodera z dekodere.....	65
Rekonstruowanie obrazów.....	67
Wizualizowanie przestrzeni ukrytej	68
Generowanie nowych obrazów	69
Autokodery wariacyjne	71
Koder.....	72
Funkcja straty.....	78
Szkolenie autokodera wariacyjnego	79
Analiza autokodera wariacyjnego	81
Eksploracja przestrzeni ukrytej	82
Zbiór danych CelebA	83
Szkolenie autokodera wariacyjnego	84
Analiza autokodera wariacyjnego	86
Generowanie nowych twarzy.....	87
Arytmetyka przestrzeni ukrytej.....	88
Morfing pomiędzy twarzami	89
Podsumowanie.....	90
4. Generatywne sieci antagonistyczne.....	91
Wprowadzenie	92
Głęboka splotowa siec GAN (DCGAN).....	94
Zbiór danych Bricks.....	94
Dyskryminator.....	95
Generator	97
Szkolenie DCGAN.....	101

Analiza DCGAN	106
Szkolenie GAN: rady i wskazówki	107
GAN Wassersteina z karą gradientową (WGAN-GP)	110
Strata Wassersteina	111
Ograniczenie Lipschitza	112
Wymuszanie ograniczenia Lipschitza	113
Kara gradientowa w stracie	114
Szkolenie sieci WGAN-GP	115
Analiza sieci WGAN-GP	117
Warunkowa GAN (CGAN)	119
Architektura sieci CGAN	119
Szkolenie sieci CGAN	121
Analiza sieci CGAN	123
Podsumowanie	124
5. Modele autoregresyjne	125
Wprowadzenie	126
Sieć z długą pamięcią krótkotrwałą (LSTM)	128
Zbiór danych Recipes	128
Praca z danymi tekstowymi	130
Tokenizacja	131
Tworzenie zbioru szkoleniowego	134
Architektura LSTM	134
Warstwa zagnieżdżeń	135
Warstwa LSTM	136
Komórka LSTM	138
Szkolenie LSTM	140
Analiza LSTM	142
Rozszerzenia rekurencyjnej sieci neuronowej (RNN)	145
Stosowe sieci rekurencyjne	145
Bramkowane jednostki rekurencyjne	147
Komórki dwukierunkowe	148
PixelCNN	149
Maskowane warstwy splotowe	149
Bloki resztkowe	151
Szkolenie PixelCNN	153
Analiza PixelCNN	154
Mieszanina dystrybuant	157
Podsumowanie	159
6. Modele przepływów normalizujących	161
Wprowadzenie	162
Przepływy normalizujące	164

Zamiana zmiennych	164
Wyznacznik Jacobiego	166
Równanie zamiany zmiennych	167
RealNVP	168
Zbiór danych Two Moons	168
Warstwy sprzęgające	169
Szkolenie modelu RealNVP	174
Analiza modelu RealNVP	177
Inne modele przepływu normalizującego	179
GLOW	179
FFJORD	180
Podsumowanie	181
7. Modele oparte o energię	183
Wprowadzenie	184
Modele oparte o energię	185
Zbiór danych MNIST	186
Funkcja energii	187
Próbkowanie z użyciem dynamiki Langevina	189
Szkolenie z dywergencją przeciwną	191
Analiza modeli opartych o energię	195
Inne modele oparte o energię	196
Podsumowanie	197
8. Modele dyfuzji	199
Wprowadzenie	200
Odszumiające modele dyfuzji (DDM)	202
Zbiór danych Flowers	202
Proces dyfuzji w przód	204
Sztuczka reparametryzacyjna	205
Rozkłady dyfuzji	206
Proces dyfuzji odwrotnej	208
Odszumiający model U-Net	211
Szkolenie modelu dyfuzji	218
Próbkowanie z odszumiającego modelu dyfuzji	219
Analiza modelu dyfuzji	222
Podsumowanie	225
 Część III Zastosowania	
9. Transformery	229
Wprowadzenie	230
GPT	230

Zbiór danych z Wine Reviews	231
Uwaga	232
Zapytania, klucze i wartości	233
Uwaga wielogłowicowa	235
Nieformalne maskowanie	236
Blok transformerów	238
Kodowanie pozycyjne	241
Szkolenie GPT	243
Analiza GPT	245
Inny transformer	247
T5	248
GPT-3 i GPT-4	251
ChatGPT	252
Podsumowanie	256
10. Zaawansowane GAN	257
Wprowadzenie	258
ProGAN	258
Szkolenie progresywne	259
Wyniki	265
StyleGAN	265
Sieć mapująca	267
Sieć syntetyzująca	267
Wyniki z StyleGAN	268
StyleGAN2	268
Modulacja i demodulacja wagi	270
Regularyzacja długości ścieżki	271
Brak progresywnego wzrostu	272
Wyniki z StyleGAN2	273
Inne ważne odmiany GAN	274
GAN (SAGAN) skupiony na sobie	274
BigGAN	276
VQ-GAN	277
ViT VQ-GAN	280
Podsumowanie	282
11. Generowanie muzyki	283
Wprowadzenie	284
Transformery do generowania muzyki	285
Zbiór danych suita na wiolonczelę Bacha	286
Parsowanie plików MIDI	286
Tokenizacja	288
Tworzenie zbioru szkoleniowego	289

Sinusoidalne kodowanie pozycji	290
Wiele wejść i wyjść	292
Analiza transformera generującego muzykę	294
Tokenizacja muzyki polifonicznej	298
MuseGAN	302
Zbiór danych z chorałem Bacha	302
Generator MuseGAN	305
Krytyk MuseGAN	311
Analiza MuseGAN	312
Podsumowanie	314
12. Modele światowe	317
Wprowadzenie	318
Uczenie przez wzmacnianie	318
Środowisko CarRacing	320
Ogólne spojrzenie na model światowy	322
Architektura	322
Szkolenie	325
Zbieranie losowo zbieranych danych	325
Szkolenie VAE	327
Architektura VAE	327
Analiza VAE	329
Zbieranie danych do wyszkolenia MDN-RNN	331
Szkolenie MDN-RNN	331
Architektura MDN-RNN	331
Próbkowanie z MDN-RNN	332
Szkolenie sterownika	333
Architektura sterownika	333
CMA-ES	334
Zrównoleglenie CMA-ES	336
Szkolenie In-Dream	338
Podsumowanie	341
13. Modele multimodalne	343
Wprowadzenie	344
DALL.E 2	345
Architektura	346
Koder tekstu	346
CLIP	346
Model prior	351
Dekoder	353
Przykłady z DALL.E 2	356
Imagen	360

Architektura	360
DrawBench	361
Przykłady z Imagen	362
Stable Diffusion	363
Architektura	363
Przykłady ze Stable Diffusion	364
Flamingo	364
Architektura	365
Vision Encoder	365
Perceiver Resampler	366
Language Model	368
Przykłady z Flamingo	371
Podsumowanie	372
14. Wnioski	375
Oś czasu generatywnej sztucznej inteligencji	376
2014–2017: Era VAE i GAN	378
2018–2019: Era transformera	378
2020–2022: Era dużego modelu	379
Obecny stan generatywnej sztucznej inteligencji	380
Duże modele językowe	380
Modele generowania kodu na podstawie tekstu	384
Modele generowania obrazów na podstawie tekstu	385
Inne zastosowania	389
Przyszłość generatywnej sztucznej inteligencji	390
Generatywna sztuczna inteligencja w życiu codziennym	391
Generatywna sztuczna inteligencja w miejscu pracy	392
Generatywna sztuczna inteligencja w edukacji	394
Etyka i wyzwania generatywnej sztucznej inteligencji	395
Końcowe przemyślenia	397
Indeks	399
O autorze	411

Przedmowa

Ta książka staje się częścią mojego życia. Po znalezieniu egzemplarza w moim salonie, zapytałem syna: „Kiedy ją dostałeś?”. Odpowiedział: „Wtedy, kiedy mi ją dałeś”, zdumiony moimi problemami z pamięcią. Przeglądając wspólnie różne rozdziały, doszedłem do wniosku, że *Generatywne uczenie głębokie* jest odpowiednikiem *Gray's Anatomia*¹ dla generatywnej sztucznej inteligencji.

Autor z niesamowitą przejrzystością i podbudowującym autorytetem analizuje anatomię generatywnej sztucznej inteligencji. Przedstawia naprawdę niezwykłą relację z szybko zmieniającą się dziedziną, opartą na pragmatycznych przykładach, wciągających narracjach i odniesieniach tak aktualnych, że czyta się ją jak żywą historię.

Poprzez swoje dekonstrukcje autor utrzymuje poczucie zachwytu i podekscytowania potencjałem generatywnej sztucznej inteligencji – co jest szczególnie widoczne w fascynującym zwieńczeniu książki. Po obnażeniu technologii przypomina nam, że stoimy u progu nowej ery inteligencji, epoki, w której generatywna sztuczna inteligencja stanowi zwierciadło naszego języka, naszej sztuki i naszej kreatywności; odzwierciedlając nie tylko to, co stworzyliśmy, ale także to, co moglibyśmy stworzyć – co możemy stworzyć – ograniczone jedynie „naszą własną wyobraźnią”. Główny motyw modeli generatywnych w sztucznej inteligencji głęboko we mnie rozbrzmiewa, ponieważ widzę dokładnie te same motywy pojawiające się w naukach przyrodniczych – mianowicie postrzeganie siebie jako generatywnych modeli świata, w którym żyjemy. Podejrzewam, że w kolejnej edycji tej książki przeczytamy o splocie sztucznej i naturalnej inteligencji. Do tego czasu na mojej półce z książkami będę trzymać to wydanie obok mojego egzemplarza *Gray Anatomia* i innych skarbów.

– Karl Friston, członek Towarzystwa Królewskiego,
profesor neuronauki University College London

1 *Anatomia* autorstwa Henry'ego Graya to jeden z najbardziej znanych na świecie podręczników dla studentów medycyny (przyp. tłum.).

Wstęp

Czego stworzyć nie mogę, tego nie pojmuję.

–Richard Feynman

Generatywna sztuczna inteligencja to jedna z najbardziej rewolucyjnych technologii naszych czasów, zmieniająca sposób, w jaki komunikujemy się z maszynami. Jej potencjał zrewolucjonizowania sposobu, w jaki żyjemy, pracujemy i bawimy się, był przedmiotem niezliczonych rozmów, debat i przewidywań. Ale co by było, gdyby w tej potężnej technologii tkwił jeszcze większy potencjał? A co, jeśli możliwości generatywnej sztucznej inteligencji wykraczają poza nasze aktualne wyobrażenie? Przyszłość generatywnej sztucznej inteligencji może być bardziej ekscytująca, niż kiedykolwiek sądziliśmy... Od samego początku poszukiwaliśmy możliwości tworzenia oryginalnych i pięknych dzieł. W przypadku wczesnych ludzi przybierało to formę malowideł jaskiniowych przedstawiających dzikie zwierzęta i abstrakcyjne wzory, tworzonych przy użyciu pigmentów ostrożnie i metodycznie umieszczanych na skałach. Era romantyzmu dała nam mistrzostwo symfonii Czajkowskiego, z ich zdolnością do wzbudzania uczuć triumfu i tragedii poprzez fale dźwiękowe, splecione ze sobą, tworzące wspaniałą muzykę i harmonię. Ostatnio zdarza się nam gnać o północy do księgarni, żeby kupić opowiadania o fikcyjnym czarodzieju, ponieważ kombinacja liter tworzy narrację, która zmusza nas do przewrócenia strony i dowiedzenia się, co przydarzy się naszemu bohaterowi.

Nic więc dziwnego, że ludzkość zaczęła zadawać ostateczne pytanie dotyczące kreatywności: czy możemy stworzyć coś, co samo w sobie jest twórcze?

Oto pytanie, na które stara się odpowiedzieć generatywna sztuczna inteligencja. Dzięki ostatniemu postępowi metodologii i technologii jesteśmy obecnie w stanie zbudować maszyny, które potrafią malować oryginalne dzieła sztuki w określonym stylu, pisać spójne bloki tekstu o długofalowej strukturze, komponować przyjemną muzykę do słuchania i opracowywać zwycięskie strategie dla złożonych gier poprzez generowanie wymagowanych scenariuszy przyszłości. To dopiero początek generatywnej rewolucji, która nie pozostawi nas bez innego wyboru, jak tylko znaleźć odpowiedzi na niektóre z najważniejszych pytań dotyczących mechaniki kreatywności i ostatecznie tego, co to znaczy być człowiekiem.

Krótko mówiąc, nigdy nie było lepszego momentu na poznanie generatywnej sztucznej inteligencji – zatem zaczynamy!

Cel i podejście

Książka nie wymaga wcześniejszej wiedzy na temat generatywnej sztucznej inteligencji. Wszystkie kluczowe koncepcje wprowadzimy od podstaw w sposób intuicyjny i przystępny, więc nie martw się, jeśli nie masz doświadczenia z generatywną sztuczną inteligencją. Jesteś we właściwym miejscu!

Zamiast tylko opisywać techniki, które są obecnie w modzie, książka służy jako kompletny przewodnik po modelowaniu generatywnym, obejmujący szeroki zakres rodzin modeli. Nie ma jednej techniki, która byłaby obiektywnie *lepsz*a czy *gorsza* od innych – w rzeczywistości wiele najnowocześniejszych modeli łączy obecnie pomysły z szerokiego spektrum podejść do modelowania generatywnego. Z tego powodu ważne jest, aby być na bieżąco z rozwojem we wszystkich obszarach generatywnej sztucznej inteligencji, zamiast skupiać się na jednym konkretnym rodzaju techniki. Jedno jest pewne: dziedzina generatywnej sztucznej inteligencji rozwija się szybko i nigdy nie wiadomo, skąd nadejdzie kolejny przełomowy pomysł!

Mając na uwadze powyższe, zastosuję podejście polegające na pokazaniu, jak trenować własne modele generatywne na własnych danych, zamiast polegać na wstępnie wyszkolonych, gotowych modelach. Chociaż obecnie istnieje wiele imponujących modeli generatywnych typu open source, które można pobrać i uruchomić poprzez kilka linijek kodu, celem tej książki jest głębsze poznanie ich architektury i projektowanie na bazie podstawowych zasad, aby można było w pełni zrozumieć, w jaki sposób działają i móc zaprogramować od podstaw przykłady każdej z technik przy użyciu języka Python i Keras.

Reasumując, niniejszą książkę można traktować jako mapę obecnego krajobrazu generatywnej sztucznej inteligencji, obejmującą zarówno teorię, jak i zastosowania praktyczne, zawierającą kompletne działające przykłady kluczowych modeli z literatury. Przeanalizujemy krok po kroku każdy kod, korzystając z jasnych drogowskazów pokazujących, w jaki sposób kod implementuje teorię leżącą u podstaw poszczególnych technik. Książkę tę można czytać od deski do deski lub używać jej jako podręcznika, w który można się zagłębiać. I przede wszystkim mam nadzieję, że zostanie uznana za przydatną i przyjemną lekturę!



W całej książce znajdują się krótkie, alegoryczne historie, które pomogą w wyjaśnianiu mechaniki niektórych budowanych przez nas modeli. Wierzę, że jednym z najlepszych sposobów nauczania nowej abstrakcyjnej teorii jest najpierw przekształcenie jej w coś, co nie jest tak abstrakcyjne, jak na przykład opowieść, przed zagłębieniem się w wyjaśnienia techniczne. Historia i wyjaśnienie modelu to ta sama mechanika wyjaśniona w dwóch różnych dziedzinach – dlatego podczas poznawania szczegółów technicznych poszczególnych modeli może okazać się przydatne odniesienie do odpowiedniej historii!

Wymagania wstępne

W książce założono, że posiadasz doświadczenie w kodowaniu w Pythonie. Jeśli nie znasz języka Python, najlepiej zacząć od strony *LearnPython.org*. W internecie jest dostępnych wiele bezpłatnych zasobów, które pozwolą rozwinąć wiedzę na temat języka Python na tyle, aby można było pracować z przykładami zawartymi w niniejszej książce.

Ponadto, ponieważ niektóre modele są opisane przy użyciu notacji matematycznej, przydatna będzie solidna znajomość algebry liniowej (na przykład mnożenia macierzy) i ogólnej teorii prawdopodobieństwa. Przydatnym źródłem może okazać się swobodnie dostępna książka Deisenrotha i in. *Mathematics for Machine Learning* (Cambridge University Press).

Autor książki nie zakłada żadnej wcześniejszej wiedzy na temat modelowania generatywnego (przeanalizujemy kluczowe pojęcia w rozdziale 1) ani TensorFlow czy Keras (biblioteki te zostaną przedstawione w rozdziale 2).

Mapa drogowa

Książka została podzielona na trzy części.

Część I stanowi ogólne wprowadzenie do modelowania generatywnego i uczenia głębokiego, podczas którego zbadamy podstawowe koncepcje leżące u podstaw wszystkich technik opisanych w dalszej części książki:

- W rozdziale 1, „Modelowanie generatywne” zdefiniujemy modelowanie generatywne i rozważymy pewien uproszczony przykład, którego można użyć do zrozumienia niektórych kluczowych pojęć ważnych dla wszystkich modeli generatywnych. Przedstawimy również taksonomię rodzin modeli generatywnych, którą omówimy w części II tej książki.
- W rozdziale 2, „Uczenie głębokie” rozpoczniemy naszą eksplorację uczenia głębokiego i sieci neuronowych od zbudowania naszego pierwszego przykładu perceptronu wielowarstwowego (MLP) przy użyciu biblioteki Keras. Następnie zmodyfikujemy go, uwzględniając warstwy splotowe i inne ulepszenia, aby zaobserwować różnicę w wydajności.

Część II omawia sześć kluczowych technik, których będziemy używać do budowania modeli generatywnych, wraz z praktycznymi przykładami każdej z nich:

- W rozdziale 3 „Autokodery wariacyjne” rozważymy wariacyjny autokoder (VAE) i przekonamy się, jak można go wykorzystać do generowania obrazów twarzy i przekształceń między twarzami w przestrzeni ukrytej modelu.
- W rozdziale 4 „Generatywne sieci antagonistyczne” zbadamy generatywne sieci antagonistyczne (GAN) do generowania obrazu, w tym głębokie splotowe GAN,

warunkowe GAN i ich rozwinięcia, takie jak GAN Wassersteina, dzięki którym proces uczenia staje się bardziej stabilny.

- W rozdziale 5 „Modele autoregresyjne” zwrócimy naszą uwagę na modele autoregresyjne, zaczynając od wprowadzenia do rekurencyjnych sieci neuronowych, takich jak sieci z długą pamięcią krótkotrwałą (LSTM) do generowania tekstu i PixelCNN do generowania obrazów.
- W rozdziale 6 „Modele przepływów normalizujących” skupimy się na przepływach normalizujących, w tym na intuicyjnej teoretycznej eksploracji techniki i praktycznym przykładzie budowania modelu RealNVP w celu generowania obrazów.
- W rozdziale 7 „Modele oparte o energię” omówimy modele oparte o energię, w tym istotne metody, takie jak szkolenie przy użyciu przeciwnych rozbieżności i próbkowanie przy zastosowaniu dynamiki Langevina.
- W rozdziale 8 „Modele dyfuzji” przejdziemy przez praktyczny przewodnik po budowaniu modeli dyfuzji, które napędzają wiele najnowocześniejszych modeli generowania obrazu, takich jak DALL.E 2 i Stable Diffusion.

Na koniec, w części III na bazie tych podstaw zbadamy wewnętrzne działanie najnowocześniejszych modeli generowania obrazów, tekstów, komponowania muzyki i uczenia ze wzmocnieniem, wykorzystując model środowiska:

- W rozdziale 9 „Transformery” zbadamy pochodzenie i szczegóły techniczne modeli StyleGAN, a także innych najnowszych sieci GAN do generowania obrazu, takich jak VQ-GAN.
- W rozdziale 10 „Zaawansowane GAN” rozważymy architekturę transformera, a także znajdzie się w nim praktyczny przewodnik dotyczący tworzenia własnej wersji GPT do generowania tekstów.
- W rozdziale 11 „Generowanie muzyki” skierujemy naszą uwagę na generowanie muzyki. Znajdzie się w nim również przewodnik po pracy z danymi muzycznymi i zastosowaniu technik, takich jak transformery i MuseGAN.
- W rozdziale 12 „Modele światowe” zobaczymy, jak modele generatywne mogą być stosowane w kontekście uczenia się ze wzmocnieniem, z zastosowaniem modeli świata i metod opartych na transformerach.
- W rozdziale 13 „Modele multimodalne” wyjaśnimy wewnętrzne działanie czterech najnowocześniejszych modeli multimodalnych, które uwzględniają więcej niż jeden typ danych, w tym DALL.E 2, Imagen i Stable Diffusion do zamiany tekstu na obrazy i Flamingo, modelu języka wizualnego.
- W rozdziale 14 „Wnioski” podsumujemy najważniejsze dotychczasowe kamienie milowe generatywnej sztucznej inteligencji i omówimy, w jaki sposób w nadchodzących latach generatywna sztuczna inteligencja zrewolucjonizuje nasze codzienne życie.

Zmiany w drugim wydaniu

Dziękuję wszystkim, którzy przeczytali pierwsze wydanie tej książki – naprawdę cieszę się, że tak wielu z Was uznało ją za przydatne źródło informacji i przekazało swoje uwagi na temat rzeczy, które chcielibyście zobaczyć w drugim wydaniu. Dziedzina generatywnego uczenia głębokiego znacząco się rozwinęła od czasu opublikowania pierwszego wydania w 2019 roku, więc oprócz odświeżenia poprzedniej treści dodałem kilka nowych rozdziałów w celu dostosowania materiału do aktualnego stanu wiedzy.

Poniżej znajduje się podsumowanie głównych aktualizacji w zakresie poszczególnych rozdziałów oraz całościowego ulepszenia książki:

- Rozdział 1 zawiera teraz punkt dotyczący różnych rodzin modeli generatywnych oraz taksonomię ich powiązań.
- Rozdział 2 zawiera ulepszone diagramy i bardziej szczegółowe objaśnienia kluczowych pojęć.
- Rozdział 3 został odświeżony o nowy przykład i dodatkowe objaśnienia.
- Rozdział 4 zawiera teraz objaśnienie warunkowych architektur GAN.
- Rozdział 5 zawiera teraz punkt dotyczący modeli autoregresyjnych dla obrazów (np. PixelCNN).
- Rozdział 6 to zupełnie nowy rozdział, opisujący model RealNVP.
- Rozdział 7 to także nowy rozdział, skupiający się na technikach, takich jak dynamika Langevina i dywergencja kontrastowa.
- Rozdział 8 to na nowo napisany rozdział poświęcony odzsumianiu modeli dyfuzji, które napędzają wiele współczesnych, najnowocześniejszych zastosowań.
- Rozdział 9 stanowi rozwinięcie materiału zawartego w zakończeniu pierwszego wydania, z głębszym skupieniem się na architekturach różnych modeli StyleGAN i nowym materiale na temat VQ-GAN.
- Rozdział 10 to nowy rozdział, który szczegółowo wyjaśnia architekturę transformera.
- Rozdział 11 zawiera współczesne architektury transformerów, zastępujące modele LSTM z pierwszego wydania.
- Rozdział 12 zawiera zaktualizowane diagramy i opisy, z punktem dotyczącym tego, w jaki sposób to podejście wpływa na aktualny stan wiedzy na temat uczenia ze wzmocnieniem.
- Rozdział 13 to nowy rozdział, który szczegółowo wyjaśnia, jak działają imponujące modele, takie jak DALL.E 2, Imagen, Stable Diffusion i Flamingo.
- Rozdział 14 został zaktualizowany, aby odzwierciedlić wyjątkowy postęp w tej dziedzinie od czasu pierwszej edycji i dać pełniejszy i bardziej szczegółowy obraz kierunku, w jakim zmierza generatywna sztuczna inteligencja w przyszłości.

- Zostały uwzględnione wszelkie komentarze przekazane w ramach opinii na temat pierwszego wydania i wychwycone literówki (zgodnie z moją najlepszą wiedzą!).
- Na początku poszczególnych rozdziałów zostały dodane ich cele, dzięki czemu przed rozpoczęciem lektury można się zapoznać z kluczowymi tematami poruszonymi w rozdziale.
- Niektóre alegoryczne historie zostały przepisane, aby były bardziej zwięzłe i jasne – cieszę się, że tak wielu czytelników stwierdziło, że pomogły one lepiej zrozumieć kluczowe pojęcia!
- Nagłówki i podtytuły każdego rozdziału zostały dostosowane tak, aby było jasne, które części rozdziału skupiają się na wyjaśnieniach, a które na budowaniu własnych modeli.

Inne zasoby

Jako ogólne wprowadzenie do uczenia maszynowego i uczenia głębokiego gorąco polecam następujące książki:

Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems autorstwa Aurélien Géron (<https://learning.oreilly.com/library/view/hands-on-machine-learning/9781098125967>) (wydawnictwo O'Reilly)²

Deep Learning with Python autorstwa Francois Chollet (wydawnictwo Manning)

Źródłem większości artykułów zawartych w tej książce jest arXiv, bezpłatne repozytorium artykułów naukowych. Obecnie częstą praktyką jest publikowanie artykułów w arXiv przed ich pełną recenzją. Przeglądanie najnowszych zgłoszeń to świetny sposób, aby być na bieżąco z najnowocześniejszymi osiągnięciami w danej dziedzinie.

Gorąco polecam również witrynę *Papers with Code* (<https://paperswithcode.com/>), w której można znaleźć najświeższy stan wiedzy w zakresie rezultatów różnych zadań uczenia maszynowego, a także linki do artykułów i oficjalnych repozytoriów GitHub. Jest to doskonałe źródło informacji dla każdego, kto chce szybko zrozumieć, które techniki obecnie uzyskują najwyższe rezultaty spośród różnych zadań i z pewnością pomogło mi podjąć decyzję, które techniki uwzględnić w tej książce.

2 Wydanie polskie *Uczenie maszynowe z użyciem Scikit-Learn, Keras i TensorFlow*, Helion 2023.

Konwencje stosowane w tej książce

W tej książce zastosowano następujące konwencje typograficzne:

Kursywa

Wskazuje nowe pojęcia, adresy URL, adresy e-mail, nazwy plików i rozszerzenia plików.

Stała szerokość

Używane w przypadku poleceń i listingów programów, a także w akapitach w odniesieniu do elementów programu, takich jak nazwy zmiennych lub funkcji.

Stała szerokość kursywą

Wskazuje tekst, który należy zastąpić wartościami dostarczonymi przez użytkownika lub wartościami określonymi przez kontekst.



Ten element oznacza wskazówkę lub sugestię.



Ten element oznacza uwagę ogólną.



Ten element oznacza ostrzeżenie lub uwagę.

Baza kodu

Przykłady kodu zawarte w tej książce można znaleźć w repozytorium GitHub. Celowo zadbałem o to, aby żaden z modeli nie wymagał do trenowania zbyt dużej ilości zasobów obliczeniowych, dzięki czemu można rozpocząć trenowanie własnych modeli bez konieczności poświęcania dużej ilości czasu i wydawania pieniędzy na drogi sprzęt. W repozytorium znajduje się obszerny przewodnik dotyczący rozpoczęcia pracy z Dockerem i, jeśli to konieczne, konfiguracji zasobów chmurowych z procesorami graficznymi w Google Cloud.

W stosunku do pierwszej edycji wprowadzono następujące zmiany w bazie kodu:

- Wszystkie przykłady można teraz uruchamiać z poziomu jednego notebooka, zamiast potrzeby importowania fragmentów kodu z modułów spośród całej bazy

kodu. Dzięki temu można wykonać każdy przykład komórka po komórce i dokładnie poznać, jak zbudowany jest każdy model, kawałek po kawałku.

- Sekcje każdego notebooka są z grubsza dopasowane pomiędzy przykładami.
- Wiele przykładów w tej książce wykorzystuje fragmenty kodu z niesamowitego open source'owego repozytorium Keras – ma to na celu uniknięcie tworzenia całkowicie odrębnego repozytorium przykładów generatywnej sztucznej inteligencji wykorzystującego Keras, gdy istnieją już doskonałe implementacje dostępne w witrynie Keras. W całej tej książce oraz w repozytorium dodałem odniesienia i linki do autorów oryginalnego kodu, z którego korzystałem poprzez stronę Keras.
- W porównaniu z pierwszą edycją dodałem nowe źródła danych i usprawniłem proces zbierania danych – w tej wersji dostępny jest skrypt, który można łatwo uruchomić w celu zebrania danych z wymaganych źródeł w celu uczenia przykładów zawartych w książce, przy użyciu narzędzi takich jak Kaggle API (<https://oreil.ly/8ibPw>).

Korzystanie z przykładów kodu

Materiały dodatkowe (przykłady kodu, ćwiczenia itp.) są dostępne do pobrania pod adresem https://github.com/davidADSP/Generative_Deep_Learning_2nd_Edition.

Jeśli masz pytanie techniczne lub problem z użyciem przykładów kodu, wyślij e-mail na adres bookquestions@oreilly.com.

Ta książka ma pomóc w wykonywaniu Twojej pracy. Ogólnie rzecz biorąc, jeśli w tej książce znajduje się przykładowy kod, możesz go używać w swoich programach i dokumentacji. Nie musisz kontaktować się z nami w celu uzyskania zgody, chyba że powielasz znaczną część kodu. Dla przykładu napisanie programu korzystającego z kilku fragmentów kodu z tej książki nie wymaga pozwolenia. Sprzedaż lub dystrybucja przykładów z książek O'Reilly już tak. Odpowiadanie na pytanie poprzez cytowanie tej książki i przykładowego kodu nie wymaga pozwolenia. Wykorzystywanie znacznej ilości przykładowego kodu z tej książki do dokumentacji produktu wymaga już pozwolenia.

Doceniamy atrybucję, ale jej nie wymagamy. Uznanie autorstwa zazwyczaj obejmuje tytuł, autora, wydawcę i numer ISBN. Dla przykładu: „*Generative Deep Learning*, 2nd edition, by David Foster (O'Reilly). Copyright 2023 Applied Data Science Partners Ltd., 978-1-098-13418-1.”

Jeśli uznasz, że użycie przykładów kodu wykracza poza dozwolony użytek lub zgodę udzieloną powyżej, skontaktuj się z nami pod adresem permissions@oreilly.com.

Jak można się z nami skontaktować

Wszelkie uwagi i pytania dotyczące tej książki prosimy kierować do wydawcy:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (w Stanach Zjednoczonych i w Kanadzie)
707-829-0515 (międzynarodowo lub lokalnie)
707-829-0104 (faks)

Dla tej książki stworzyliśmy stronę internetową, na której zamieszczamy listę errat, przykładów i wszelkich dodatkowych informacji. Strona jest dostępna pod adresem <https://oreil.ly/generative-dl>.

Wyślij e-mail na bookquestions@oreilly.com, aby skomentować lub zadać pytania techniczne dotyczące tej książki.

Aktualności i informacje o naszych książkach i kursach można znaleźć na stronie <https://oreilly.com>.

Znajdź nas na LinkedIn: <https://linkedin.com/company/oreilly-media>

Śledź nas na Twitterze: <https://twitter.com/oreillymedia>

Oglądaj nas na YouTube'ie: <https://youtube.com/oreillymedia>

Podziękowania

Jest tak wiele osób, którym chciałbym podziękować za pomoc w pisaniu tej książki. Po pierwsze, chciałbym podziękować wszystkim, którzy poświęcili czas na techniczną recenzję książki – w szczególności Vishweshowi Raviemu Shrimali, Lipi Deepaakshi Patnaik, Lubie Elliot i Lornie Barclay. Dziękuję także Samirowi Bico za pomoc w przejrzaniu i przetestowaniu kodu źródłowego dołączonego do tej książki. Twój wkład był nieoceniony.

Ogromne podziękowania kieruję także do moich kolegów z Applied Data Science Partners (<https://adsp.ai/>), Ross Witeszczak, Amy Bull, Ali Parandeh, Zine Eddine, Joęgo Rowe, Gerty Salillari, Aleshii Parkes, Evelyn Kireilyte, Riccarda Tolli, Mai Do, Khaleela Syed i Willa Holmesa. Wasza cierpliwość do mnie, gdy poświęcałem czas na dokończenie książki, jest niezwykle cenna i nie mogę się doczekać wszystkich projektów związanych z uczeniem maszynowym, które wspólnie zrealizujemy w przyszłości! Szczególne podziękowania należą się Ross – gdybyśmy nie zdecydowali się na wspólne założenie firmy, ta książka mogłaby nigdy nie nabrać kształtu, dlatego dziękuję za wiarę we mnie jako partnera biznesowego!

Chcę także podziękować każdemu, kto kiedykolwiek nauczył mnie czegokolwiek z matematyki. Jestem szczęściarzem, że miałem w szkole fantastycznych nauczycieli

matematyki, którzy rozwinęli moje zainteresowanie tym przedmiotem i zachęcili mnie do kontynuowania jej zgłębiania na uniwersytecie. Chciałbym podziękować za zaangażowanie i za to, że zrobiliście wszystko, aby podzielić się ze mną swoją wiedzą na ten temat.

Ogromne podziękowania kieruję do personelu O'Reilly za przeprowadzenie mnie przez proces pisania tej książki. Specjalne podziękowania należą się Michele Cronin, która była obecna na każdym etapie, przekazując mi przydatne uwagi i wysyłając życzliwe przypomnienia, abym kontynuował uzupełnianie rozdziałów! Dziękuję także Nicole Butterfield, Christopherowi Faucherowi, Charlesowi Roumeliotisowi i Suzanne Huston za umożliwienie wydania książki oraz Mike'owi Loukidesowi za pierwszy kontakt i zapytanie, czy byłbym zainteresowany jej napisaniem. Wszyscy od początku bardzo wspieraliście ten projekt i chcę wam podziękować za udostępnienie mi platformy, na której mogę pisać o czymś, co kocham.

W trakcie całego procesu pisania moja rodzina była nieustającym źródłem otuchy i wsparcia. Ogromne podziękowania kieruję do mojej mamy, Gillian Foster, za sprawdzanie każdej linijki tekstu pod kątem literówek i przede wszystkim za nauczenie mnie sumowania! Twoja dbałość o szczegóły była niezwykle pomocna podczas korekty tej książki i jestem naprawdę wdzięczny za wszystkie możliwości, które dałaś mi zarówno ty, jak i tata. Mój tata, Clive Foster, nauczył mnie, jak programować komputer – ta książka jest pełna praktycznych przykładów, a to dzięki jego cierpliwości we wczesnych latach mojego życia, gdy jako nastolatek grzebałem w języku BASIC, próbując tworzyć gry związane z futbolem. Mój brat Rob Foster to najskromniejszy geniusz, jakiego kiedykolwiek można spotkać, zwłaszcza w dziedzinie lingwistyki – rozmowa z nim na temat sztucznej inteligencji i przyszłości uczenia maszynowego opartego na tekście okazała się niezwykle pomocna. Na koniec chciałbym podziękować mojej Nanie, która zawsze była dla nas wszystkich nieustającym źródłem inspiracji i dobrej zabawy. Jej miłość do literatury była jednym z powodów, dla których zdecydowałem, że napisanie książki będzie ekscytującym zajęciem.

Chciałbym także podziękować mojej żonie, Lornie Barclay. Oprócz bezkresnego wsparcia i oceanu filiżanek herbaty podczas całego procesu pisania, rygorystycznie i szczegółowo sprawdzałaś każde słowo tej książki. Nie dałbym rady tego dokonać bez ciebie. Dziękuję, że zawsze byłaś przy mnie i że uczyniłaś tę podróż o wiele przyjemniejszą. Obiecuję, że nie będę mówić o generatywnej sztucznej inteligencji przy stole przez przynajmniej kilka dni po wydaniu książki.

Na koniec chciałbym podziękować naszej pięknej córeczce Alinie za zapewnienie niekończącej się rozrywki podczas długich nocy spędzonych na pisaniu książek. Twoje urocze chichoty były idealnym podkładem muzycznym do mojego pisania. Dziękuję, że jesteś moją inspiracją i że zawsze trzymasz mnie w gotowości. To ty jesteś prawdziwym mózgiem tej operacji.

Wprowadzenie do generatywnego głębokiego uczenia

Część I to ogólne wprowadzenie do modelowania generatywnego i głębokiego uczenia – dwóch dziedzin, które musimy zrozumieć, aby rozpocząć generatywne głębokie uczenie!

W rozdziale 1 zdefiniujemy modelowanie generatywne i rozważymy uproszczony przykład, którego można użyć do zrozumienia niektórych kluczowych pojęć ważnych dla wszystkich modeli generatywnych. Przedstawimy również taksonomię rodzin modeli generatywnych, którą omówimy w części II tej książki.

Rozdział 2 to przewodnik po narzędziach i technikach głębokiego uczenia, które będą nam potrzebne do rozpoczęcia tworzenia bardziej złożonych modeli generatywnych. W szczególności zbudujemy nasz pierwszy przykład głębokiej sieci neuronowej – perceptron wielowarstwowy (MLP) – przy użyciu biblioteki Keras. Następnie, aby móc zaobserwować różnicę w wydajności, dostosujemy go do uwzględnienia warstw splotowych i innych rozwinięć.

Pod koniec Części I Czytelnik posiada odpowiednie rozumienie podstawowych pojęć leżących u podstaw wszystkich technik opisanych w dalszych częściach książki.

Modelowanie generatywne

Cele rozdziału

W tym rozdziale:

- Poznamy kluczowe różnice pomiędzy modelami generatywnymi i dyskryminacyjnymi.
- Poprzez prosty przykład zrozumiemy odpowiednie właściwości modelu generatywnego.
- Dowiemy się, jakie są podstawowe koncepcje probabilistyczne leżące u podstaw modeli generatywnych.
- Poznamy różne rodziny modeli generatywnych.
- Sklonujemy bazę kodu dołączoną do tej książki, co umożliwi rozpoczęcie budowania modeli generatywnych!

Rozdział ten stanowi ogólne wprowadzenie do dziedziny modelowania generatywnego.

Zacznijemy od łagodnego teoretycznego wprowadzenia do modelowania generatywnego i zobaczymy, w jaki sposób jest ono naturalnym odpowiednikiem szerzej badanego modelowania dyskryminacyjnego. Następnie ustalimy ramy opisujące pożądane właściwości, jakie powinien posiadać dobry model generatywny. Przedstawimy także podstawowe koncepcje probabilistyczne, które warto znać, aby w pełni docenić, w jaki sposób różne podejścia radzą sobie z wyzwaniem modelowania generatywnego.

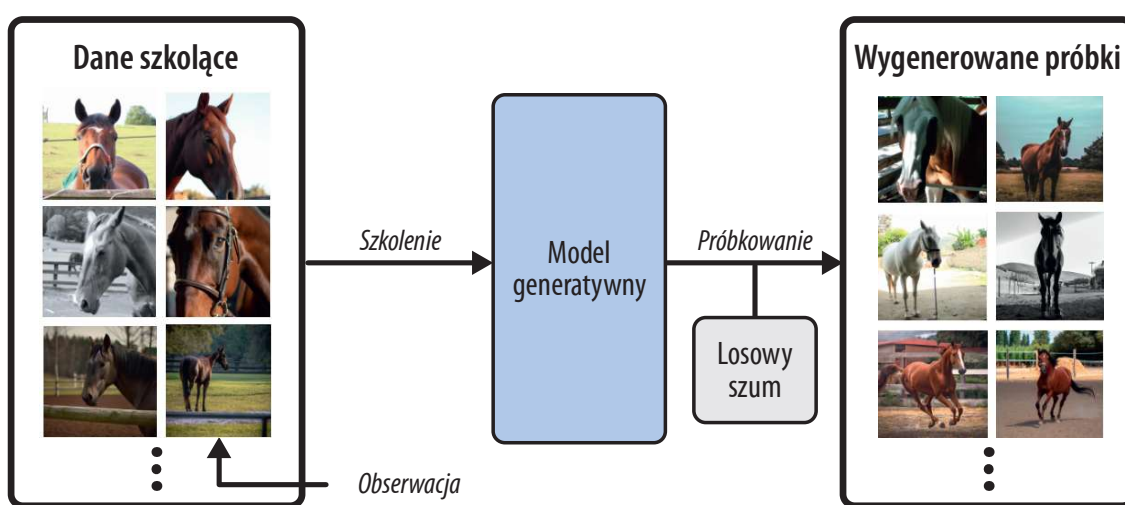
W naturalny sposób doprowadzi nas to do przedostatniej części, która przedstawia sześć szerokich rodzin modeli generatywnych dominujących aktualnie w tej dziedzinie. W ostatniej części wyjaśnimy, jak rozpocząć pracę z bazą kodu dołączoną do tej książki.

Czym jest modelowanie generatywne?

Modelowanie generatywne można w sposób ogólny zdefiniować w następujący sposób:

Modelowanie generatywne to gałąź uczenia maszynowego, która polega na wyszkoleniu modelu w celu wygenerowania nowych danych podobnych do podanego zbioru danych.

Co to oznacza w praktyce? Załóżmy, że mamy zbiór danych zawierający zdjęcia koni. Na tym zbiorze danych możemy *wyszkolić* model generatywny do uchwycenia zasad rządzących złożonymi relacjami między pikselami na obrazach koni. Następnie możemy *pobrać próbki* z tego modelu, aby stworzyć nowatorskie, realistyczne obrazy koni, które nie istniały w oryginalnym zbiorze danych. Proces ten przedstawiono na rysunku 1.1.



Rysunek 1.1 Model generatywny wyszkolony do generowania realistycznych zdjęć koni

Do zbudowania modelu generatywnego potrzebujemy zbioru danych składającego się z wielu przykładów bytu, którą staramy się wygenerować. Są one znane pod nazwą *dane szkoleniowe*, a jeden taki punkt danych jest nazywany *obserwacją*.

Każda obserwacja składa się z wielu *cech*. Dla problemu z generowaniem obrazu cechami są zwykle wartości poszczególnych pikseli. Dla problemu generowania tekstu cechami mogą być pojedyncze słowa lub grupy liter. Naszym celem jest zbudowanie modelu, który będzie w stanie wygenerować nowe zestawy cech wyglądające tak, jakby zostały utworzone przy użyciu tych samych reguł, co oryginalne dane. Pod względem koncepcyjnym, w przypadku generowania obrazu, jest to niezwykle trudne zadanie, biorąc pod uwagę ogromną liczbę sposobów przypisania wartości poszczególnym pikselom i stosunkowo niewielką liczbę takich układów, które składają się na obraz obiektu, który staramy się wygenerować.

Model generatywny musi być także *probabilistyczny*, a nie *deterministyczny*, ponieważ chcemy mieć możliwość generowania wielu różnych wariantów wyniku, zamiast uzyskiwać za każdym razem ten sam. Jeśli nasz model jest jedynie ustalonym obliczeniem, takim jak średnia wartość każdego piksela w zbiorze danych uczących, nie jest on generatywny.

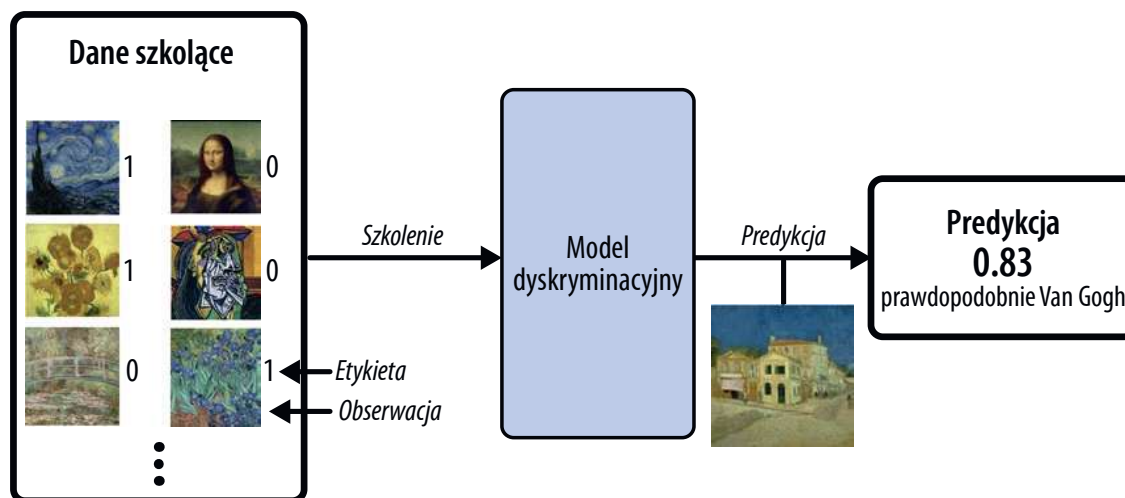
Model generatywny musi zawierać składnik losowy, który wpływa na poszczególne próbki generowane przez model.

Innymi słowy, możemy sobie wyobrazić, że istnieje nieznan rozkład prawdopodobieństwa, który wyjaśnia, dlaczego niektóre obrazy prawdopodobnie zostaną znalezione w zbiorze danych uczących, a inne nie. Naszym zadaniem jest zbudowanie modelu, który możliwie najdokładniej naśladuje ten rozkład, a następnie pobranie z niego próbek w celu wygenerowania nowych, odrębnych obserwacji, które wyglądają tak, jakby mogły zostać uwzględnione w oryginalnym zbiorze uczącym.

Modelowanie generatywne vs. dyskryminacyjne

Aby naprawdę zrozumieć, co ma na celu osiągnięcie modelowanie generatywne i dlaczego jest to ważne, warto porównać je z jego ekwiwalentem, *modelowaniem dyskryminacyjnym*. Jeśli ktoś studiował uczenie maszynowe, większość problemów, na które napotka, najprawdopodobniej będzie miało charakter dyskryminacyjny. Zapoznajmy się z przykładem, który zobrazuje nam tę różnicę.

Załóżmy, że mamy zbiór danych zawierający obrazy, niektóre namalowane przez Van Gogha, pozostałe przez innych artystów. Mając wystarczającą ilość danych, moglibyśmy wyszkolić model dyskryminacyjny, aby określić, czy dany obraz został namalowany przez Van Gogha. Nasz model dowiedziałby się, że określone kolory, kształty i tekstury z większym prawdopodobieństwem wskazują, że obraz jest dziełem holenderskiego mistrza, a w przypadku obrazów posiadających te cechy model odpowiednio wzmocniłby swoje przewidywania. Rysunek 1.2 przedstawia proces modelowania dyskryminacyjnego – należy zwrócić uwagę, czym różni się on od procesu modelowania generatywnego pokazanego na rysunku 1.1.



Rysunek 1.2 Model dyskryminacyjny wyszkolony do predykacji, czy dany obraz został namalowany przez Van Gogha

Podczas wykonywania modelowania dyskryminacyjnego każda obserwacja w danych uczących ma *etykietę*. W przypadku binarnego problemu klasyfikacji, takiego jak

dyskryminator artysty, obrazy Van Gogha zostaną oznaczone jako 1, a obrazy innych artystów niż Van Gogh – jako 0. Nasz model nauczy się następnie, jak rozróżniać te dwie grupy i wskaże prawdopodobieństwo, że nowa obserwacja ma etykietę 1 – tj. że został namalowany przez Van Gogha.

W przeciwieństwie do tego modelowanie generatywne nie wymaga etykietowania zbioru danych, ponieważ zajmuje się generowaniem zupełnie nowych obrazów, a nie próbą predykcji etykiety danego obrazu.

Zdefiniujmy formalnie tego typu modelowanie, posługując się notacją matematyczną:

Modelowanie dyskryminacyjne szacuje $p(y|\mathbf{x})$

Oznacza to, że modelowanie dyskryminacyjne ma na celu zamodelowanie prawdopodobieństwa etykiety y przy pewnej obserwacji \mathbf{x} .

Modelowanie generatywne szacuje $p(\mathbf{x})$

Oznacza to, że modelowanie generatywne ma na celu zamodelowanie prawdopodobieństwa zaobserwowania obserwacji \mathbf{x} . Próbkowanie z tego rozkładu pozwala nam wygenerować nowe obserwacje.



Warunkowe modele generatywne

Należy zwrócić uwagę, że możemy również zbudować model generatywny do modelowania prawdopodobieństwa warunkowego $p(\mathbf{x}|y)$ – prawdopodobieństwa zaobserwowania obserwacji \mathbf{x} z określoną etykietą y .

Dla przykładu, jeśli nasz zbiór danych zawiera różne rodzaje owoców, możemy nakazać naszemu modelowi generatywnemu, aby generował konkretnie obraz jabłka.

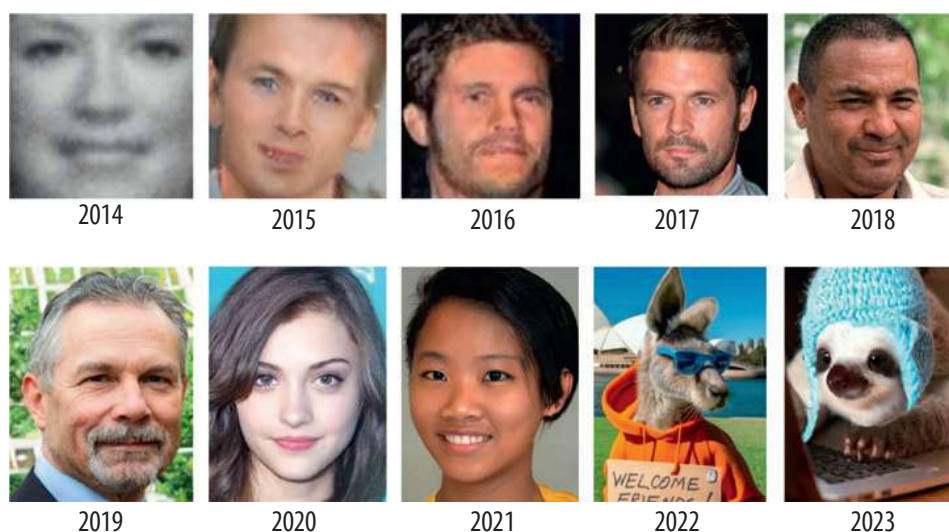
Należy zauważyć, że nawet gdybyśmy byli w stanie zbudować doskonały model dyskryminacyjny do identyfikacji obrazów Van Gogha, nadal nie mielibyśmy pojęcia, jak stworzyć obraz wyglądający jak dzieło Van Gogha. Może jedynie generować prawdopodobieństwa w odniesieniu do istniejących obrazów, ponieważ model do tego został wyszkolony. Zamiast tego moglibyśmy wyszkolić model generatywny i próbować ten model w celu wygenerowania obrazów, które mają duże szanse na przynależność do oryginalnego zbioru danych treningowych.

Powstanie modelowania generatywnego

Do niedawna modelowanie dyskryminacyjne było siłą napędową większości postępów w uczeniu maszynowym. Dzieje się tak dlatego, że w przypadku każdego problemu dyskryminacyjnego odpowiadający mu problem modelowania generatywnego jest zwykle znacznie trudniejszy do rozwiązania. Dla przykładu znacznie łatwiej jest wyszkolić model, aby przewidywał, czy obraz jest dziełem Van Gogha, niż wyszkolić model do wygenerowania od zera obrazu w stylu Van Gogha.

Podobnie znacznie łatwiej jest wyszkolić model, aby przewidywał, czy strona tekstu została napisana przez Charlesa Dickensa, niż zbudować model generujący zbiór akapitów w stylu Dickensa. Do niedawna większość problemów generatywnych była zwyczajnie poza zasięgiem i wielu wątpiło, czy kiedykolwiek uda się je rozwiązać. Kreatywność uznawano za zdolność czysto ludzką, z którą nie mogła rywalizować sztuczna inteligencja.

Jednak w miarę dojrzewania technologii uczenia maszynowego założenie to stopniowo słabnie. W ciągu ostatnich 10 lat wiele z najciekawszych osiągnięć w tej dziedzinie nastąpiło dzięki nowatorskim zastosowaniom uczenia maszynowego do zadań modelowania generatywnego. Dla przykładu rysunek 1.3 pokazuje uderzający postęp, jaki od 2014 roku nastąpił w generowaniu obrazu twarzy.



Rysunek 1.3 W ciągu ostatniej dekady generowanie twarzy przy użyciu modelowania generatywnego znacznie się poprawiło (na podstawie Brundage et al., 2018)¹

Z modelowaniem dyskryminacyjnym jest nie tylko łatwiej się uporać, ale także w przeszłości było łatwiej dostosowywane do praktycznych problemów w różnych branżach, niż modelowanie generatywne. Na przykład lekarz może odnieść korzyść z modelu, który przewiduje, czy dany obraz siatkówki wykazuje oznaki jaskry, ale niekoniecznie skorzysta z modelu, który może wygenerować nowe obrazy tylnej części oka.

Jednak to również zaczyna się zmieniać wraz z rozprzestrzenianiem się firm oferujących usługi generatywne, które są ukierunkowane na konkretne problemy biznesowe. Dla przykładu można teraz uzyskać dostęp do interfejsów API, które generują oryginalne wpisy dla bloga o określonej tematyce, tworzą różnorodne zdjęcia produktu w dowolnym ustawieniu lub piszą treści do mediów społecznościowych i teksty reklam pasujące do danej marki i przekazu docelowego. Istnieją również wyraźnie pozytywne zastosowania generatywnej sztucznej inteligencji w branżach, takich jak projektowanie

¹ Miles Brundage et al., „The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation,” February 20, 2018, https://www.eff.org/files/2018/02/20/malicious_ai_report_final.pdf.

gier i kinematografia, gdzie modele przeszkolone w zakresie tworzenia wideo i muzyki zaczynają wносить wartość dodaną.

Modelowanie generatywne i sztuczna inteligencja

Oprócz praktycznych zastosowań modelowania generatywnego (z których wiele jeszcze nie zostało odkrytych) istnieją trzy głębsze powody, dla których modelowanie generatywne można uznać za klucz do odblokowania znacznie bardziej wyrafinowanej formy sztucznej inteligencji, która wykracza poza to, co może osiągnąć samo modelowanie dyskryminacyjne.

Po pierwsze, z czysto teoretycznego punktu widzenia, nie powinniśmy ograniczać naszego uczenia maszynowego do prostego kategoryzowania danych. Dla kompletności, powinniśmy również zająć się modelami uczącymi, które oddadzą pełniejsze zrozumienie rozkładu danych, poza jakimiś konkretnymi etykietami. Jest to niewątpliwie problem trudniejszy do rozwiązania, ze względu na dużą wymiarowość przestrzeni możliwych wyników oraz stosunkowo niewielką liczbę wytworów, które zaklasyfikowalibyśmy jako należące do zbioru danych. Jednakże, jak się przekonamy, wiele z tych samych technik, które napędzają rozwój modelowania dyskryminacyjnego, takich jak głębokie uczenie, może być również wykorzystywanych w modelach generatywnych.

Po drugie jak zobaczymy w rozdziale 12, modelowanie generatywne jest obecnie wykorzystywane do napędzania postępu w innych obszarach sztucznej inteligencji, takich jak uczenie ze wzmocnieniem (badanie agentów uczących w celu optymalizacji celu w środowisku metodą prób i błędów). Załóżmy, że chcemy wyszkolić robota, aby poruszał się po wskazanym terenie. Tradycyjne podejście polegałoby na przeprowadzeniu wielu eksperymentów, podczas których agent wypróbuje różne strategie w terenie, lub komputerowej symulacji terenu. Z biegiem czasu agent nauczyłby się, które strategie są skuteczniejsze od innych i dlatego stopniowo by się ulepszał. Wyzwaniem związanym z tym podejściem jest to, że jest ono dość nieelastyczne, ponieważ jest szkolone w zakresie optymalizacji polityki dla jednego konkretnego zadania. Alternatywnym podejściem, które ostatnio zyskało na popularności, jest wyszkolenie agenta w zakresie uczenia się *modelu świata* środowiska przy użyciu modelu generatywnego, niezależnego od konkretnego zadania. Agent może szybko dostosować się do nowych zadań, testując strategie we własnym modelu świata, a nie w rzeczywistym środowisku, które często jest wydajniejsze obliczeniowo i nie wymaga ponownego uczenia od podstaw dla każdego nowego zadania.

I wreszcie, jeśli mamy szczerze powiedzieć, że zbudowaliśmy maszynę, która uzyskała inteligencję porównywalną z ludzką, modelowanie generatywne z pewnością musi być częścią rozwiązania. Jednym z najwspanialszych przykładów modelu generatywnego w świecie przyrody jest osoba czytająca tę książkę. Poświęćmy chwilę i zastanówmy się, jakim niesamowitym modelem generatywnym jesteśmy. Można zamknąć oczy i wyobrazić sobie, jak wyglądałby słoń pod każdym możliwym kątem. Można wyobrazić sobie wiele różnych, prawdopodobnych zakończeń swojego ulubionego programu telewizyjnego i zaplanować nadchodzący tydzień, wyobrażając sobie różne wersje przyszłości i podejmując odpowiednie działania. Obecna teoria neuronaukowa sugeruje, że nasze

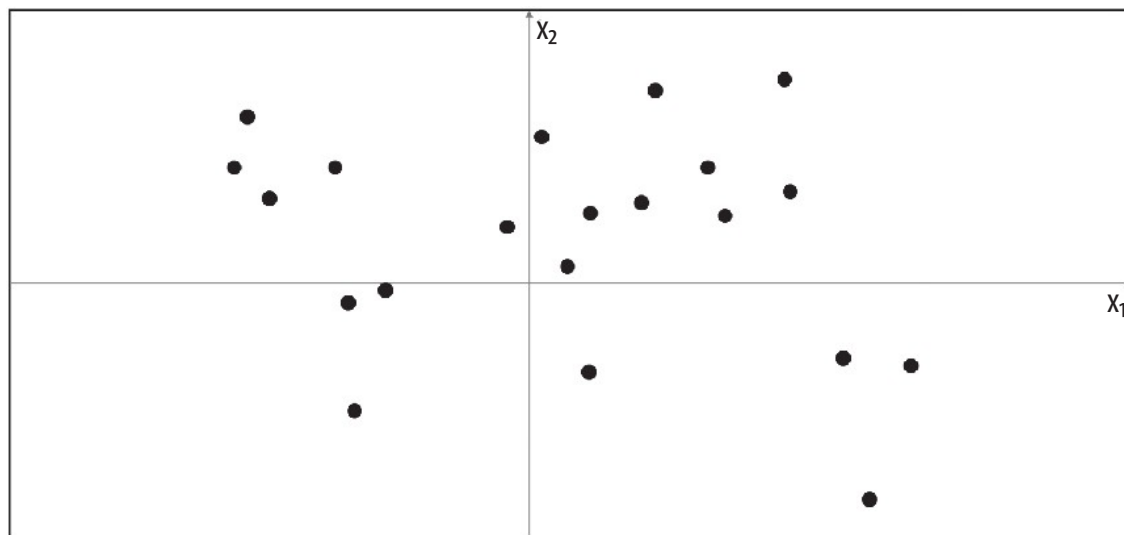
postrzeganie rzeczywistości nie jest bardzo złożonym modelem dyskryminującym, działającym na podstawie naszych bodźców zmysłowych w celu wytworzenia przewidywań tego, czego doświadczamy, ale zamiast tego jest modelem generatywnym, który jest trenowany od urodzenia w celu stworzenia symulacji naszego otoczenia, które dokładnie odpowiadają przyszłości. Niektóre teorie sugerują nawet, że wynikiem tego modelu generatywnego jest to, co bezpośrednio postrzegamy jako rzeczywistość. Jest oczywiste, że głębokie zrozumienie tego, w jaki sposób możemy zbudować maszyny umożliwiające nabycie tej zdolności, będzie miało kluczowe znaczenie dla naszego dalszego zrozumienia działania mózgu i ogólnej sztucznej inteligencji.

Nasz pierwszy model generatywny

Mając to na uwadze, rozpocznijmy naszą podróż do ekscytującego świata modelowania generatywnego. Na początek przyjrzymy się uproszczonemu przykładowi modelu generatywnego i przedstawimy kilka pomysłów, które pomogą nam pracować nad bardziej złożonymi architekturami, z którymi zetkniemy się w dalszej części książki.

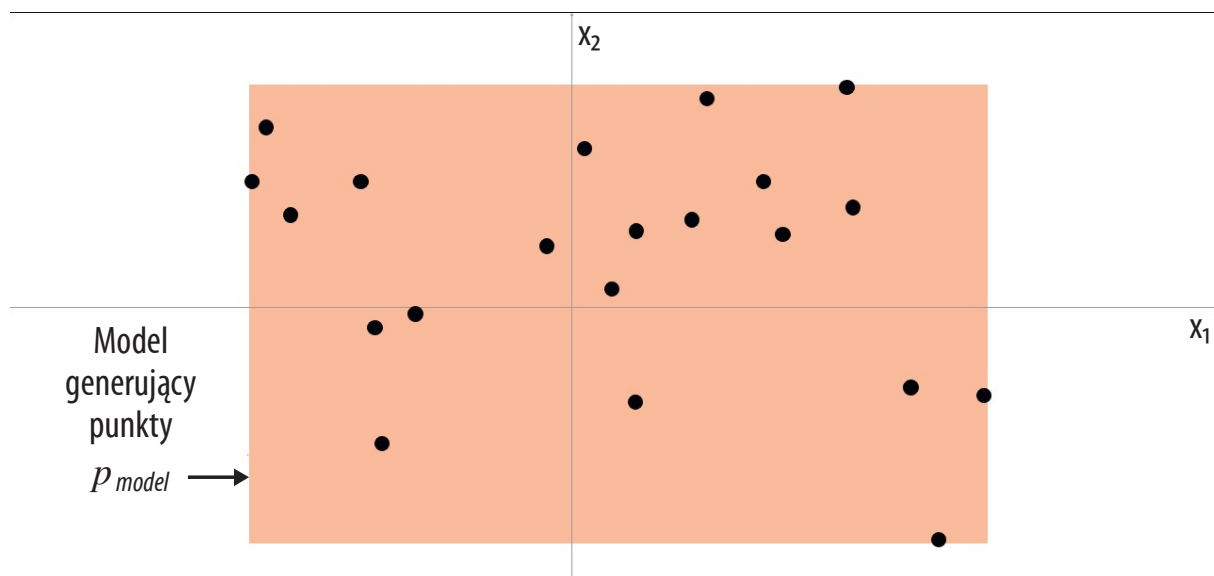
Witaj, świecie!

Zacznijmy od zagrania w grę polegającą na modelowaniu generatywnym w zaledwie dwóch wymiarach. Wybrałem regułę, która posłużyła do wygenerowania zbioru punktów \mathbf{X} przedstawionego na rysunku 1.4. Nazwijmy tę regułę p_{data} . Twoim wyzwaniem jest wybranie innego punktu $\mathbf{x} = (x_1, x_2)$ w przestrzeni, który ma wyglądać, jakby został wygenerowany na podstawie tej samej reguły.



Rysunek 1.4 Zbiór punktów w dwóch wymiarach, wygenerowany przez nieznaną regułę p_{data}

Gdzie je wybierzesz? Prawdopodobnie wykorzystasz swoją wiedzę na temat istniejących punktów danych do skonstruowania modelu mentalnego, p_{model} – miejsc w przestrzeni, w której istnieje większe prawdopodobieństwo, że powinien znaleźć się punkt. Pod tym względem p_{model} jest estymacją p_{data} . Być może zdecydowałeś, że p_{model} powinien wyglądać jak na rysunku 1.5 – prostokątna ramka, w której mogą znaleźć się punkty, i obszar poza ramką, w którym nie ma szans na znalezienie się jakichkolwiek punktów.



Rysunek 1.5 Pomarańczowe pole, p_{model} to oszacowanie rzeczywistego rozkładu generującego dane – p_{data}

W celu wygenerowania nowej obserwacji, można wybrać losowo punkt w ramce lub, bardziej formalnie, próbkę z rozkładu p_{model} . Gratulacje, właśnie udało Ci się zbudować swój pierwszy model generatywny! Zostały przez Ciebie wykorzystane dane treningowe (czarne punkty) do skonstruowania modelu (pomarańczowy obszar), z którego można łatwo pobierać próbki w celu wygenerowania innych punktów, które wydają się należeć do zbioru treningowego.

Sformalizujemy teraz to myślenie w ramy, które pomogą nam zrozumieć, co próbuje osiągnąć modelowanie generatywne.

Ramy modelowania generatywnego

Nasze motywacje i cele związane z budowaniem modelu generatywnego możemy ująć w następujący sposób.

Ramy modelowania generatywnego

- Mamy zbiór danych z obserwacjami X .
- Zakładamy, że obserwacje zostały wygenerowane według jakiegoś nieznanego rozkładu p_{data} .
- Chcemy zbudować model generatywny p_{model} , który naśladuje p_{data} . Jeśli osiągniemy ten cel, będziemy mogli pobierać próbki z p_{model} , aby wygenerować obserwacje, które wyglądają na pochodzące z p_{data} .
- Dlatego pożądanymi właściwościami p_{model} są:

Dokładność

Jeśli dla wygenerowanej obserwacji p_{model} jest wysoki, powinien wyglądać tak, jakby został uzyskany z p_{data} . Jeśli dla wygenerowanej obserwacji p_{model} jest niski, nie powinien wyglądać, jakby został uzyskany z p_{data} .

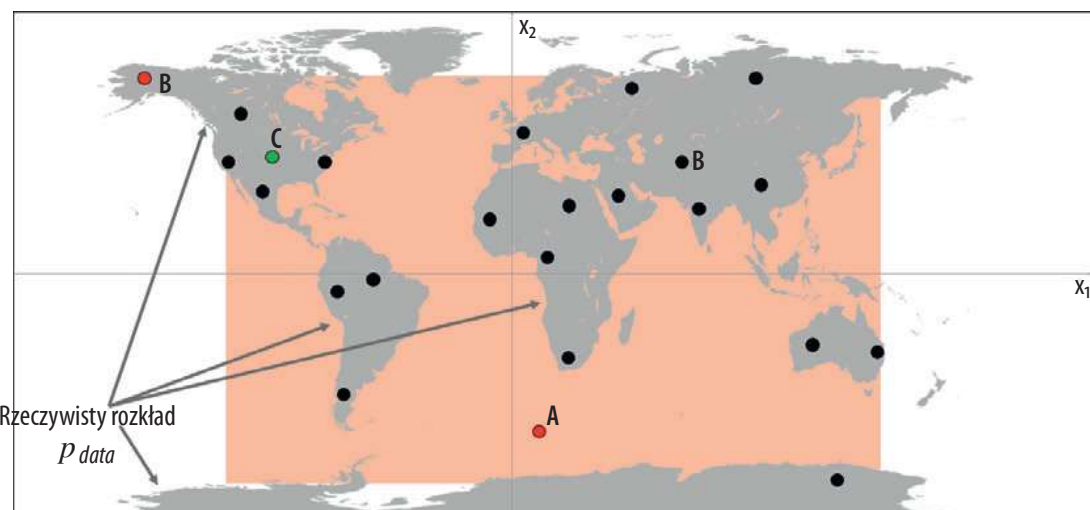
Generacja

Powinno być możliwe łatwe pobranie próbki nowej obserwacji z p_{model} .

Reprezentacja

Powinno być możliwe zrozumienie, jak p_{model} reprezentuje różne wysokopoziomowe cechy w danych.

Ujawnijmy teraz p_{data} , rzeczywisty rozkład generujący dane, i przekonajmy się, jak te ramy mają zastosowanie w tym przykładzie. Jak widać na rysunku 1.6, regułą generowania danych jest po prostu rozkład jednostajny na powierzchni lądów świata, bez szans na wskazanie punktu na morzu.



Rysunek 1.6 Pomarańczowe pole p_{model} to oszacowanie rzeczywistego rozkładu generującego dane p_{data} (szary obszar)

Oczywiście, nasz model p_{model} jest nadmiernym uproszczeniem p_{data} . Możemy zbadać punkty A, B i C, aby zrozumieć sukcesy i porażki naszego modelu względem tego, jak dokładnie naśladuje on p_{data} :

- Punkt A to obserwacja wygenerowana przez nasz model, ale wydaje się, że nie została wygenerowana przez p_{data} , ponieważ znajduje się na morzu.
- Punkt B nigdy nie mógłby zostać wygenerowany przez p_{model} , ponieważ znajduje się poza pomarańczowym prostokątem. Dlatego nasz model ma pewne luki w zdolności do generowania obserwacji w całym zakresie potencjalnych możliwości.
- Punkt C jest obserwacją, która może zostać wygenerowana przez p_{model} , a także przez p_{data} .

Pomimo swoich niedociągnięć, model jest łatwy do próbkowania, ponieważ jest to zwykle rozkład jednostajny w pomarańczowym polu. W trakcie próbkowania możemy łatwo wybrać losowo punkt z jego wnętrza.

Z całą pewnością możemy też powiedzieć, że nasz model jest prostą reprezentacją leżącego u jego podstaw złożonego rozkładu, i który uwzględnia niektóre z jego wysokopoziomowych cech. Prawdziwy rozkład jest podzielony na obszary z dużą ilością obszarów lądowych (kontynenty) i te bez obszarów lądowych (morza). Jest to wysokopoziomowa cecha, która dotyczy również naszego modelu, z tą różnicą, że mamy jeden duży kontynent, a nie wiele.

W tym przykładzie zademonstrowano podstawowe pojęcia stojące za modelowaniem generatywnym. Problemy, którymi zajmiemy się w tej książce, będą znacznie bardziej złożone i wielowymiarowe, ale fundamentalne ramy, poprzez które będziemy podchodzić do problemu, pozostaną niezmiennie.

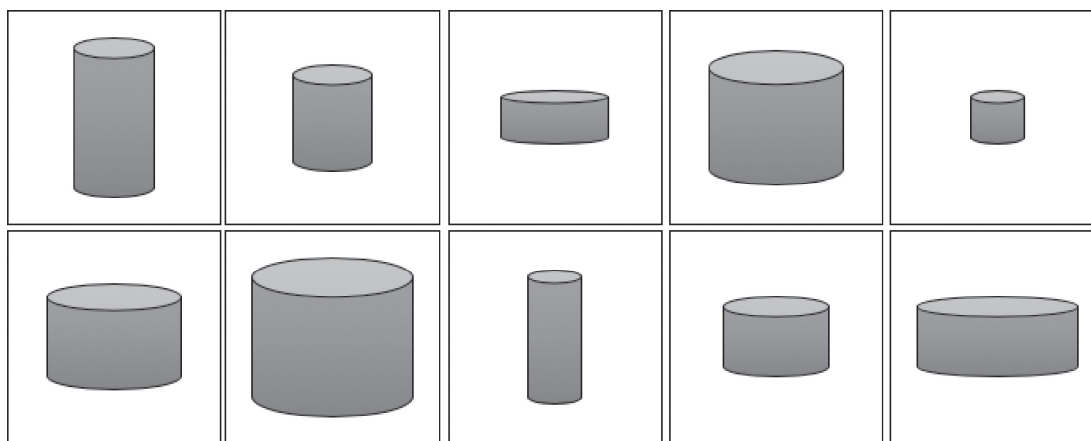
Uczenie reprezentacji

Warto nieco bardziej zagłębić się w to, co mamy na myśli, mówiąc o *reprezentacji* danych wielowymiarowych, ponieważ temat będzie powracał w tej książce.

Załóżmy, że chcemy opisać swój wygląd komuś, kto szuka nas w tłumie ludzi i nie wie, jak wyglądamy. Nie zaczynamy od podania koloru pierwszego piksela na naszym zdjęciu, następnie piksela 2, potem piksela 3, itd. Zamiast tego przyjąlibyśmy rozsądne założenie, że druga osoba ma ogólne pojęcie o tym, jak wygląda przeciętny człowiek, a następnie zmienilibyśmy tę linię bazową, dodając cechy opisujące grupy pikseli, jak na przykład: *Mam bardzo jasne włosy* lub *Noszę okulary*. Mając nie więcej, niż 10 takich stwierdzeń, taka osoba będzie w stanie odwzorować opis z powrotem w postaci pikseli, aby wygenerować w swojej głowie nasz obraz. Może nie byłby idealny, ale na tyle podobny do naszego rzeczywistego wizerunku, że można było nas znaleźć wśród prawdopodobnie setek innych osób, nawet jeśli ktoś nigdy wcześniej by nas nie widział.

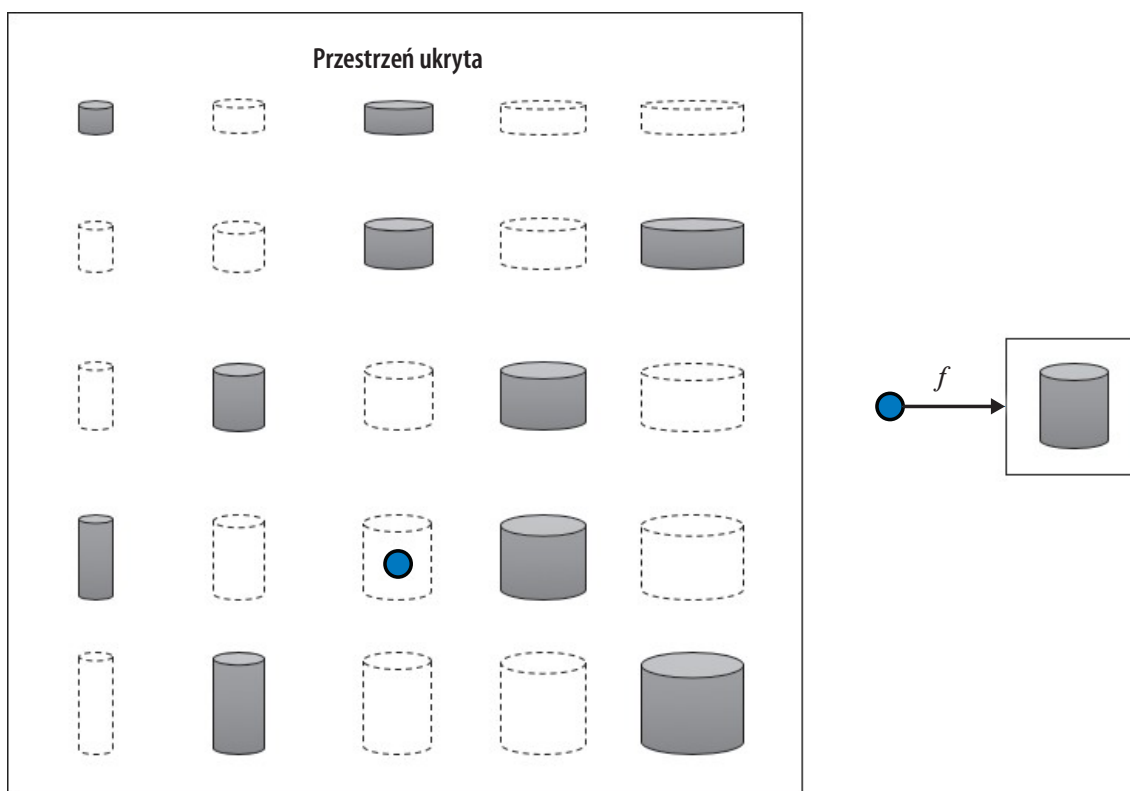
To jest główna idea stojąca za *uczeniem reprezentacji*. Zamiast starać się bezpośrednio modelować wielowymiarową przestrzeń próbek, opisujemy każdą obserwację w zbiorze uczącym przy użyciu niskowymiarowej *przestrzeni ukrytej*, a następnie uczymy się funkcji mapującej, która może pobrać punkt z przestrzeni ukrytej i odwzorować go na punkt w pierwotnej domenie. Innymi słowy, każdy punkt w przestrzeni ukrytej jest *reprezentacją* jakiejś wielowymiarowej obserwacji.

Co to oznacza w praktyce? Załóżmy, że mamy zbiór uczący złożony z obrazów szarych, cieniowanych puszek na ciastka (rysunek 1.7).



Rysunek 1.7 Zbiór danych dotyczących puszek na ciastka

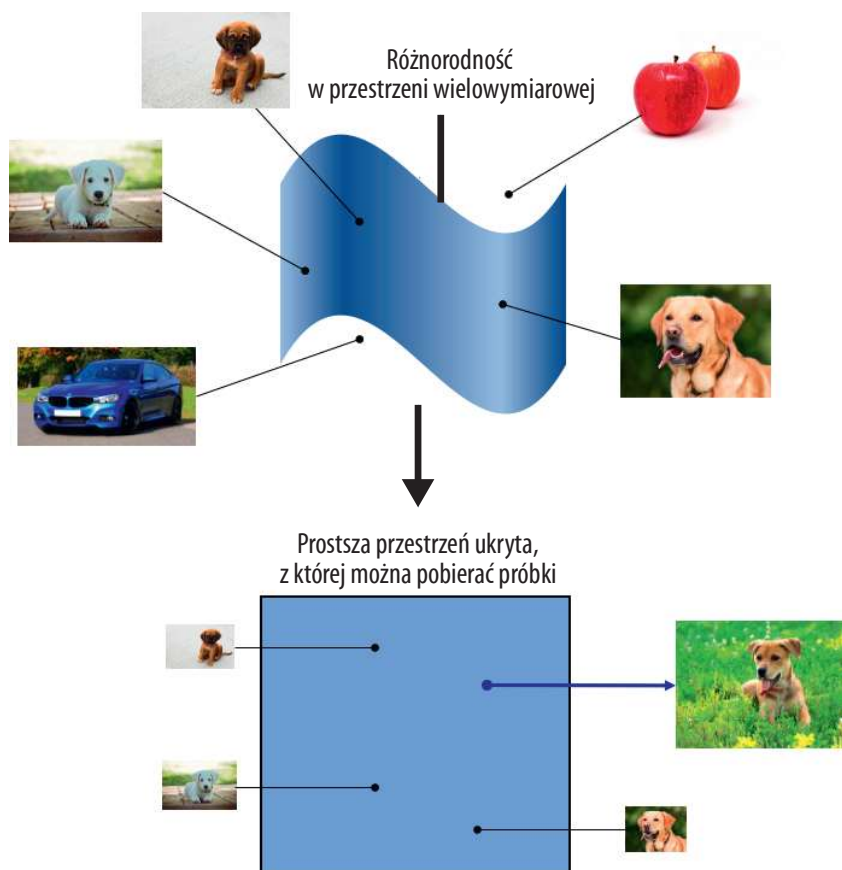
Dla nas jest oczywiste, że każdą z tych puszek wyróżniają dwie cechy: wysokość i szerokość puszki. Oznacza to, że możemy przekonwertować każdy obraz puszki na punkt w przestrzeni ukrytej o zaledwie dwóch wymiarach, mimo że zbiór uczący obrazów jest podany w wielowymiarowej przestrzeni pikseli. W szczególności oznacza to, że możemy również wygenerować obrazy puszek, które nie istnieją w zbiorze uczącym, stosując odpowiednią funkcję mapującą f do nowego punktu w przestrzeni ukrytej, jak pokazano na rysunku 1.8. Zrozumienie, że oryginalny zbiór danych można opisać prostszą przestrzenią ukrytą, nie jest takie proste dla maszyny – musiałaby najpierw ustalić, że wysokość i szerokość to dwa wymiary przestrzeni ukrytej, które najlepiej opisują ten zbiór danych, a następnie nauczyć się funkcji mapującej f , która może wybierać punkt w tej przestrzeni i odwzorować go na obraz cieniowanej szarej puszki po ciastkach. Uczenie maszynowe (w szczególności głębokie uczenie) daje nam możliwość uczenia maszyn, przez co potrafią znajdować te złożone relacje bez udziału człowieka.



Rysunek 1.8 Dwuwymiarowa ukryta przestrzeń dla puszek na ciastka i funkcja f , która odwzorowuje punkt w przestrzeni ukrytej z powrotem na oryginalną dziedzinę obrazu

Jedną z zalet modeli uczących wykorzystujących przestrzeń ukrytą jest to, że możemy wykonywać operacje, które wpływają na wysokopoziomowe właściwości obrazu, manipulując jego wektorem reprezentacji w łatwiejszej do zarządzania przestrzeni ukrytej. Przykładowo – nie jest oczywiste, jak dostosować cieniowanie każdego pojedynczego piksela, aby obraz puszek po ciastkach był wyższy. Jednak w przestrzeni ukrytej wystarczy zwiększyć ukryty wymiar *wysokość*, a następnie zastosować funkcję mapującą, aby powrócić do dziedziny obrazu. Jasny przykład tego zobaczymy w następnym rozdziale, zastosowany nie do puszek na ciastka, ale do twarzy.

Koncepcja kodowania uczącego zbioru danych w przestrzeni ukrytej, abyśmy mogli z niego pobrać próbki i zdekodować punkt z powrotem do dziedziny pierwotnej, jest powszechna w wielu technikach modelowania generatywnego, o czym przekonamy się w dalszych rozdziałach tej książki. Z matematycznego punktu widzenia techniki *kodowania-dekodowania* próbują przekształcić wysoce nieliniową *rozmaitość*, w której znajdują się dane (np. w przestrzeni pikseli), w prostszą przestrzeń ukrytą, z której można pobierać próbki, przez co staje się prawdopodobne, że dowolny punkt z przestrzeni ukrytej jest reprezentacją dobrze zbudowanego obrazu, jak pokazano na rysunku 1.9.



Rysunek 1.9 *Różnorodność psów w wielowymiarowej przestrzeni pikseli jest mapowana na prostszą przestrzeń ukrytą, z której można pobierać próbki*

Podstawowa teoria prawdopodobieństwa

Przekonaliśmy się już, że modelowanie generatywne jest ściśle powiązane ze statystycznym modelowaniem rozkładów prawdopodobieństwa. Dlatego sensowne jest teraz wprowadzenie kilku podstawowych pojęć probabilistycznych i statystycznych, które będą używane w całej książce w celu wyjaśnienia teoretycznych podstaw każdego modelu.

Jeśli problematyka prawdopodobieństwa i statystyki jest Ci obca, nie martw się. Do zbudowania wielu modeli głębokiego uczenia, które zobaczymy w dalszej części tej książki, nie jest konieczne dogłębne zrozumienie teorii statystycznej. Aby jednak móc w pełni docenić zadanie, przed którym stoimy, warto spróbować zbudować solidne zrozumienie podstawowej teorii prawdopodobieństwa. W ten sposób będziemy mieć podstawy do zrozumienia różnych rodzin modeli generatywnych, które zostaną przedstawione w dalszej części tego rozdziału.

W pierwszym kroku zdefiniujemy pięć kluczowych terminów, łącząc każdy z nich z naszym wcześniejszym przykładem modelu generatywnego, który modeluje mapę świata w dwóch wymiarach: