

O'REILLY®

2ª Edição
Atualizada com
a TensorFlow 2

Mãos à Obra: Aprendizado de Máquina com Scikit-Learn, Keras & TensorFlow

CONCEITOS, FERRAMENTAS E
TÉCNICAS PARA A CONSTRUÇÃO
DE SISTEMAS INTELIGENTES

Aurélien Géron



ALTA BOOKS
GRUPO EDITORIAL
Rio de Janeiro, 2021

Mãos à Obra: Aprendizado de Máquina com Scikit-Learn, Keras e TensorFlow – 2ª Edição Atualizada com a TensorFlow2

Copyright © 2021 da Starlin Alta Editora e Consultoria Eireli.
ISBN: 978-85-5081-548-0

Translated from original Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Copyright © 2019 Aurélien Géron. ISBN 978-1-492-03264-9. This translation is published and sold by permission of O'Reilly Media, Inc., the owner of all rights to publish and sell the same. PORTUGUESE language edition published by Starlin Alta Editora e Consultoria Eireli, Copyright © 2021 by Starlin Alta Editora e Consultoria Eireli.

Todos os direitos estão reservados e protegidos por Lei. Nenhuma parte deste livro, sem autorização prévia por escrito da editora, poderá ser reproduzida ou transmitida. A violação dos Direitos Autorais é crime estabelecido na Lei nº 9.610/98 e com punição de acordo com o artigo 184 do Código Penal.

A editora não se responsabiliza pelo conteúdo da obra, formulada exclusivamente pelo(s) autor(es).

Marcas Registradas: Todos os termos mencionados e reconhecidos como Marca Registrada e/ou Comercial são de responsabilidade de seus proprietários. A editora informa não estar associada a nenhum produto e/ou fornecedor apresentado no livro.

Impresso no Brasil — 1ª Edição, 2021 — Edição revisada conforme o Acordo Ortográfico da Língua Portuguesa de 2009.

Erratas e arquivos de apoio: No site da editora relatamos, com a devida correção, qualquer erro encontrado em nossos livros, bem como disponibilizamos arquivos de apoio se aplicáveis à obra em questão.

Accesse o site www.altabooks.com.br e procure pelo título do livro desejado para ter acesso às erratas, aos arquivos de apoio e/ou a outros conteúdos aplicáveis à obra.

Suporte Técnico: A obra é comercializada na forma em que está, sem direito a suporte técnico ou orientação pessoal/exclusiva ao leitor.

A editora não se responsabiliza pela manutenção, atualização e idioma dos sites referidos pelos autores nesta obra.

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

G377m	Géron, Aurélien
Mãos à Obra: Aprendizado de Máquina com Scikit-Learn, Keras e TensorFlow – 2ª Edição Atualizada com a TensorFlow2: Conceitos, Ferramentas e Técnicas para a Construção de Sistemas Inteligentes / Aurélien Géron ; traduzido por Cibelle Ravaglia. – Rio de Janeiro : Alta Books, 2021. 640 p. : il. ; 17cm x 24cm.	
Tradução de: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Inclui índice e apêndice. ISBN: 978-85-5081-548-0	
1. Ciência da Computação. 2. Programação de computadores. 3. Aprendizado de Máquina. 4. Sistemas Inteligentes. I. Ravaglia, Cibelle. II. Título.	
2021-2470	CDD 005.133 CDU 004.43

Elaborado por Vagner Rodolfo da Silva - CRB-89418



Rua Viúva Cláudio, 291 — Bairro Industrial do Jacaré
CEP: 20.970-031 — Rio de Janeiro (RJ)
Tels.: (21) 3278-8069 / 3278-8419
www.altabooks.com.br — altabooks@altabooks.com.br

Produção Editorial
Editora Alta Books

Gerência Comercial
Daniele Fonseca

Editor de Aquisição
José Rugeri
acquisition@altabooks.com.br

Produtores Editoriais
Illysbelle Trajano
Mária de Lourdes Borges
Thié Alves

Marketing Editorial
Livia Carvalho
Gabriela Carvalho
Thiago Brito
marketing@altabooks.com.br

Equipe de Design
Larissa Lima
Marcelli Ferreira
Paulo Gomes

Diretor Editorial
Anderson Vieira

Coordenação Financeira
Solange Souza

Produtor da Obra
Thales Silva

Equipe Ass. Editorial
Brenda Rodrigues
Caroline David
Luana Rodrigues
Mariana Portugal
Raquel Porto

Equipe Comercial
Adriana Baricelli
Daiana Costa
Fillipe Amorim
Kaique Luiz
Victor Hugo Moraes
Viviane Paiva

Atuaram na edição desta obra:

Tradução/Copidesque
Cibelle Ravaglia

Diagramação
Lucia Quaresma

Revisão Gramatical
Fernanda Lutfi
Hellen Suzuki

📧 **Ouvidoria:** ouvidoria@altabooks.com.br

Editora afiliada à:



alabr
ASSOCIAÇÃO BRASILEIRA DE
DIREITOS REPROGRÁFICOS

ASSOCIADO



Prefácio	xvii
----------------	------

PARTE I: Os conceitos básicos do aprendizado de máquina	1
1. O Cenário do Aprendizado de Máquina	3
O que É Aprendizado de Máquina?	3
Por que Usar o Aprendizado de Máquina?	4
Exemplos de Aplicações	6
Tipos de Sistemas do Aprendizado de Máquina	7
Aprendizado Supervisionado/Não Supervisionado	8
Aprendizado em batch e online	13
Aprendizado baseado em instâncias versus aprendizado baseado em modelo	15
Principais Desafios do Aprendizado de Máquina	20
Quantidade insuficiente de dados de treinamento	20
Dados de Treinamento Não Representativos	20
Dados de baixa qualidade	21
Características Irrelevantes	22
Sobreajuste dos Dados de Treinamento	23
Subajuste dos dados de treinamento	24
Um passo atrás	25
Teste e Validação	25
Ajuste de hiperparâmetro e seleção de modelo	26
Incompatibilidade de Dados	26
Exercícios	28
2. Projeto de Aprendizado de Máquina Ponta a Ponta	29
Analisando o Panorama Geral	30
Abordar o problema	31
Escolha uma medida de desempenho	32
Verifique as hipóteses	34
Obtenha os Dados	34
Criando o workspace	34
Faça o download dos dados	37
Uma rápida olhada na estrutura dos dados	38
Criando um conjunto de testes	41
Explore e Visualize os Dados para Obter Informações Úteis	44
Buscando Correlações	46
Testando combinações de atributo	48
Limpendo os dados	50
Customize os Transformadores	54

Transformação de Pipelines	56
Escolha e treine um modelo	57
Aperfeiçoe Seu Modelo	61
Grid search	61
Randomized search	63
Métodos ensemble	63
Implemente, Monitore e Faça a Manutenção de Seu Sistema	65
Teste!	68
Exercícios	68
3. Classificação.....	69
MNIST	69
Treinando um Classificador Binário	71
Cálculo de Desempenho	71
Calculando a acurácia com a validação cruzada	72
Matriz de confusão	73
Precisão e revocação	74
Trade-off precisão/revocação	75
A curva ROC	78
Classificação Multiclasse	80
Análise de Erro	83
Classificação Multirrótulo	85
Classificação Multioutput	86
Exercícios	87
4. Treinando Modelos	89
Regressão Linear	90
Equação normal	91
Complexidade computacional	93
Gradiente Descendente	94
Gradiente descendente em batch (em lote)	96
Gradiente descendente em mini-batch	101
Curvas de Aprendizado	103
Modelos Lineares Regularizados	107
Regressão de Ridge	107
Regressão de lasso	109
Elastic net	111
Early stopping	111
Regressão Logística	113
Estimando as probabilidades	113
Treinamento e função de custo	114
Fronteiras de decisão	115
Regressão softmax	117
Exercícios	120
5. Máquinas de Vetores de Suporte.....	121
Classificação Linear da SVM	121

Classificação de margem suave	122
Classificação SVM Não Linear	123
Características de similaridade	125
Kernel gaussiano RBF	126
Complexidade computacional	127
Regressão SVM	128
Por Dentro dos Bastidores	129
Função de decisão e predições	130
Objetivo do treinamento	130
Programação quadrática	131
O problema da dualidade	132
SVMs kernalizadas	133
SVMs online	135
Exercícios	136
6. Árvores de Decisão.....	137
Treinando e Visualizando uma Árvore de Decisão	137
Fazendo Predições	138
Complexidade Computacional	141
Coeficiente de Gini ou Entropia?	141
Hiperparâmetros de Regularização	142
Regressão	143
Instabilidade	145
Exercícios	146
7. Aprendizado Ensemble e Florestas Aleatórias	147
Classificadores de Votação	147
Bagging e Pasting	149
Bagging e pasting na Scikit-Learn	150
Avaliação out-of-bag (OOB)	151
Florestas aleatórias	153
Árvores extras	154
Importância da característica	154
Boosting	155
AdaBoost	155
Gradiente boosting	158
Stacking	161
Exercícios	164
8. Redução de Dimensionalidade.....	165
A Maldição da Dimensionalidade	165
As Principais Abordagens para a Redução de Dimensionalidade	166
Projeção	167
Aprendizado manifold	168
PCA	169
Preservando a variância	169
Componentes principais	170

Projetando as dimensões-d	171
Usando a Scikit-Learn	171
Taxa de variância explicada	172
Escolhendo o número adequado de dimensões	172
PCA e Compactação	173
PCA randomizada	174
PCA incremental	174
Kernel PCA	175
Selecionando um kernel e ajustando os hiperparâmetros	175
LLE	177
Outras Técnicas de Redução de Dimensionalidade	179
Exercícios	180
9. Técnicas de Aprendizado Não Supervisionado.....	181
Clusterização	182
K-Means	183
Limites do K-Means	191
Usando a clusterização para a segmentação de imagens	191
Usando a clusterização para o pré-processamento	192
Usando a clusterização para aprendizado semissupervisionado	194
DBSCAN	196
Outros algoritmos de clusterização	199
Mistura de gaussianas	200
Detecção de anomalias usando as misturas de gaussianas	204
Selecionando o número de clusters	205
Modelos de mistura de gaussianas bayesianos	208
Outros algoritmos para detecção de anomalias e novidades	211
Exercícios	212

PARTE II: Redes Neurais e Aprendizado Profundo **213**

10. Introdução às Redes Neurais Artificiais com a Biblioteca Keras	215
Dos Neurônios Biológicos aos Neurônios Artificiais	215
Neurônios biológicos	216
Cálculos lógicos com neurônios	217
A perceptron	218
Perceptron multicamadas e retropropagação	221
Regressão com MLPs	224
Classificação com MLPs	225
Implementando as MLPs com a Keras	227
Instalando a TensorFlow 2	228
Construindo um classificador de imagem usando a sequencial API	228
Construindo uma MLP de regressão usando a sequencial API	236
Construindo modelos complexos usando a funcional API	237
Usando a subclassing API para construir modelos dinâmicos	241
Salvando e armazenando um modelo	242

Usando as funções de callbacks	243
Usando o TensorBoard para visualização	244
Aperfeiçoando os Hiperparâmetros das Redes Neurais	247
Número de camadas ocultas	250
Número de neurônios por camada oculta	250
Taxa de aprendizado, tamanho do batch e outros hiperparâmetros	251
Exercícios	253
11. Treinando Redes Neurais Profundas.....	255
Problemas de Gradientes de Fuga e Explosão de Gradientes	255
Inicialização de Glorot e inicialização He	256
Funções de ativação de não saturação	258
Normalização em batch	261
Clipping do gradiente	266
Reutilizando Camadas Pré-treinadas	266
Aprendizado por Transferência com a Keras	267
Pré-treinamento não supervisionado	269
Pré-treinamento em uma tarefa auxiliar	270
Otimizadores Mais Rápidos	271
Otimização momentum	271
Gradiente acelerado de Nesterov	272
AdaGrad	273
RMSProp	274
Otimização Adam e Nadam	274
Cronograma da taxa de aprendizado	277
Evitando o Sobreajuste na Regularização	281
Regularização ℓ_1 e ℓ_2	281
Dropout	282
Dropout de Monte Carlo (MC)	284
Regularização Max-Norm	286
Diretrizes Práticas	286
Exercícios	288
12. Modelos Customizados e Treinamento com a Biblioteca TensorFlow	289
Um Tour Rápido pela TensorFlow	289
Usando a TensorFlow como NumPy	291
Tensores e operações	292
Tensores e NumPy	293
Conversões de tipo	294
Variáveis	294
Outras estruturas de dados	295
Customizando Modelos e Algoritmos de Treinamento	296
Funções de perda customizadas	296
Salvando e carregando modelos com componentes customizados	296
Funções de ativação customizadas, inicializadores, regularizadores e restrições	298
Métricas customizadas	299
Camadas customizadas	301

Modelos customizados	304
Perdas e métricas baseadas nos componentes internos do modelo	306
Calculando os gradientes usando o autodiff	307
Loops de treinamento customizados	310
Funções e Grafos da TensorFlow	313
AutoGraph e tracing	314
Regras das funções da TensorFlow	315
Exercícios	317
13. Carregando e Pré-processando Dados com a TensorFlow	319
Data API	320
Transformações de encadeamento	320
Embaralhando os dados (Shuffle)	321
Pré-processamento dos dados	324
Juntando as coisas	325
Pré-busca (prefetch)	326
Usando o conjunto de dados com a tf.keras	327
O Formato TFRecord	328
Arquivos TFRecord compactados	329
Uma breve introdução aos protobufs	329
Protobufs da TensorFlow	330
Exemplos de carregar e analisar sintagmaticamente os arquivos	331
Manipulando listas de listas com o protobuf SequenceExample	332
Pré-processando as características de entrada	333
Codificando as características categóricas com vetores one-hot	334
Codificando características categóricas com embeddings	336
Camadas de pré-processamento Keras	339
TF Transform	341
Conjuntos de Dados TensorFlow (TensorFlow Datasets)	342
Exercícios	343
14. Visão Computacional Detalhada das Redes Neurais Convolucionais	345
A Arquitetura do Córtex Visual	345
Camadas Convolucionais	347
Filtros	348
Empilhando múltiplos mapas de características	349
Implementação da TensorFlow	350
Requisitos de memória	352
Camadas de Pooling	353
Implementação da TensorFlow	354
Arquiteturas das CNNs	356
LeNet-5	357
AlexNet	358
GoogLeNet	360
VGGNet	363
ResNet	363
Xception	365

SENet	367
Implementando uma CNN ResNet-34 Usando a Keras	368
Usando Modelos Pré-treinados da Keras	369
Modelos Pré-treinados para Aprendizado por Transferência	371
Classificação e Localização	373
Detecção de Objetos	374
Redes totalmente convolucionais (FCNs)	375
You only look once (YOLO)	377
Segmentação Semântica	380
Exercícios	383
15. Processamento de Sequências Usando RNNs e CNNs	385
Neurônios Recorrentes e Camadas	385
Células de memória	387
Sequências de entrada e saída	388
Treinando as RNNs	389
Previsão de uma Série Temporal	389
Métricas de baseline	390
Implementando uma RNN simples	391
RNNs profundas	392
Previsão de diversos intervalos de tempo futuros	393
Lidando com Sequências Longas	396
O problema dos gradientes instáveis	396
Lidando com o problema de memória de curto prazo	398
Exercícios	404
16. Processamento de Linguagem Natural com RNNs e Mecanismos de Atenção	405
Gerando Texto Shakesperiano com uma Char-RNN	406
Criando o conjunto de dados de treinamento	406
Como dividir um conjunto de dados sequencial	407
Decompondo o conjunto de dados sequencial em várias janelas	408
Construindo e treinando o modelo Char-RNN	409
Usando o modelo Char-RNN	410
Gerando um falso texto shakespeariano	410
RNN stateful	411
Análise de Sentimentos	413
Mascaramento	416
Reutilizando embeddings pré-treinados	418
Rede Codificador-Decodificador para Tradução Automática Neural	419
RNNs bidirecionais	422
Heurística beam search	422
Mecanismos de Atenção	424
Mecanismo de atenção visual	426
Mecanismo de atenção, tudo que você precisa: a arquitetura Transformer	427
Inovações Recentes nos Modelos de Linguagem	434
Exercícios	436

17. Aprendizado de Representação e Aprendizado Gerativo com Autoencoders e GANs ...	437
Representações de Dados Eficientes	438
Executando uma PCA com um Autoencoder Linear Subcompleto	439
Autoencoders Empilhados	440
Implementando um autoencoder empilhado com a Keras	440
Visualizando as reconstruções	441
Visualizando o conjunto de dados Fashion MNIST	442
Pré-treinamento não supervisionado com autoencoders empilhados	443
Amarrar os pesos	444
Treinando um autoencoder de cada vez	445
Autoencoders Convolucionais	446
Autoencoders Recorrentes	446
Denoising Autoencoders	447
Autoencoders Esparsos	449
Autoencoders Variacionais	451
Gerando imagens com o Fashion MNIST	454
Redes Adversárias Gerativas	455
Os desafios de treinar as GANs	458
GANs convolucionais profundas	460
Crescimento progressivo das GANs	462
StyleGANs	464
Exercícios	466
18. Aprendizado por Reforço.....	467
Aprendendo a Otimizar Recompensas	467
Pesquisa de Políticas	469
Introdução ao OpenAI Gym	470
Políticas de Rede Neural	473
Avaliação de Ações: O Problema da Atribuição de Crédito	474
Gradientes de Política	475
Processos de Decisão de Markov	479
Aprendizado por Diferença Temporal	482
Q-learning	483
Política de exploração	485
Q-learning aproximado e deep Q-learning	485
Implementando o Deep Q-learning	486
Variantes do Deep Q-learning	490
Alvos fixos do Q-valor	490
Double DQN	490
Repetição de experiência priorizada	491
Dueling DQN	492
A Biblioteca TF-Agents	492
Instalando a TF-Agents	493
Ambientes da TF-Agents	493
Especificações do ambiente	494
Wrappers de ambiente e pré-processamento Atari	495

Arquitetura de treinamento	498
Criando a deep Q-Network	499
Criando o agente DQN	500
Criando o replay buffer e o observador correspondente	501
Criando métricas de treinamento	503
Criando o driver de coleta	504
Criando o conjunto de dados	505
Criando o ciclo de treinamento	507
Visão Geral de Alguns Algoritmos RL Populares	508
Exercícios	510
19. Treinamento e Implementação de Modelos TensorFlow em Larga Escala	511
Modelo TensorFlow Serving	511
Usando o TensorFlow Serving	512
Criando um serviço de predição no GCP AI Platform	519
Usando o serviço de predição	522
Implementando um Modelo em um Dispositivo Móvel ou Embarcado	524
Usando GPUs para Acelerar os Cálculos	527
Comprando a própria GPU	528
Usando uma máquina virtual com GPU	530
Colaboratory	531
Gerenciando a RAM da GPU	532
Colocando operações e variáveis em dispositivos	534
Execução paralela em múltiplos dispositivos	535
Treinando Modelos em Múltiplos Dispositivos	537
Paralelismo de modelo	537
Paralelismo de dados	539
Treinamento em larga escala com a distribution strategies API	543
Treinando um modelo em um cluster TensorFlow	544
Rodando jobs grandes de treinamento no Google Cloud AI Platform	547
Ajuste de hiperparâmetros black box na AI Platform	548
Exercícios	549
Obrigado!	550
A. Solução dos Exercícios	551
B. Checklist de Projeto de Aprendizado de Máquina	579
C. Problema de Dualidade SVM	583
D. Autodiff.....	585
E. Outras Arquiteturas ANN Populares.....	591
F. Estruturas Especiais de Dados	597
G. Grafos da TensorFlow	603
Índice	611