CAMBRIDGE
Cambridge University Press & Assessment
978-1-009-12330-3 — A Hands-On Introduction to Machine Learning
Chirag Shah
Index
More Information

Index

A/B testing, 3, 328, 337–340, 345
accountability (machine learning), 356–358
accuracy
  association rule, 128–129
  machine learning modeling criteria, 111
ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT), 352
adversarial learning, 344
agglomerative clustering
  defined, 186
  dendogram with maximum clustering, 187
  dendogram with minimum clustering, 32, 188
  steps to create, 186
Akaike Information Criterion (AIC), 195, 196, 200, 334–335
algorithmic accountability, 356–358
algorithmic bias, 353–354
AlphaZero, 304
Amazon Web Services (AWS)
  account set-up, 66–68
  versus Microsoft Azure, 66
  purpose of, 66
Amazon Web Services (AWS) Cloud9
  create Cloud9 environment, 71
  interface, 70
  pricing, 73
  work in Cloud9 environment, 71–72
Amazon Web Services (AWS) connect
  PuTTY, 68–69
  SSH session, 69
Anaconda
  installing and configuring, 383–384
  Python language access, 28
anomaly detection
  approaches for, 164
  defined, 166
  process of, 164–165
  uses for, 164
arithmetic operators (Python), 30
artificial intelligence (AI)
  defined, 20
  differences with machine learning, 10, 19
  responsible, 352–353
artificial neural networks (ANN)
  copying human brain interactions, 227
  versus human neurons, 228–229
  human versus artificial neurons, 228–229
  learning by example, 228
  supervised, 228
ASIMO robot, 306
association rule, 128–129
autoencoder network
  architecture, 243–244
  defined, 243, 260
  unsupervised learning example, 244–246
axon, 228, 260

backward propagation methods
  defined, 237
  process of, 237
bandits (reinforcement learning)
  contextual, 316–317
  exploration–exploitation dilemma, 315–316
  multi-arm, 316–322
  one-armed, 315
batch gradient descent
  in action, 102
  defined, 114
  example, 104–109
  methodology, 20, 102
Bayes' theorem, 382
Bayesian classification
  defined, 154
  Thompson sampling, 317
Bayesian Information Criterion (BIC), 195, 196, 200, 335
Beauty AI, 348–349
bias (machine learning)
  versus fairness, 355–356
  pervasiveness, 354
  Simpson's paradox, 355
bias, statistical, 90
bias, systemic, 353–354
bias types (machine learning)
  algorithmic/systemic, 353–354
  cognitive, 354
  data, 354
bidirectional encoder representations from transformers (BERT), 287–297
Boolean values (Python), 30
bootstrap sampling technique (random forest), 133

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-009-12330-3 — A Hands-On Introduction to Machine Learning
Chirag Shah
Index
More Information

categorical variable, 145, 367
central tendency
  defined, 372
  mean, 372–373
  median, 373
  mode, 373–374
classification. *See also* decision tree
  context of, 118
  deep learning, 272–278
  defined, 9
  k-nearest neighbors (kNN), 119–123
  logistic regression, 145–152
  naïve Bayes, 154–158
  neural networks example, 240–242
  as process of supervised learning, 118
  R language, 388–389, 392–393
  softmax regression, 152–153
  support vector machine (SVM), 159–164
classification rule, 45, 127–128
cloud computing
  defined, 49–50
  essential nature of, 50
  Google Cloud Platform (GCP), 50–59
  Microsoft Azure, 59–66
cloud platform
  certification, 74
  defined, 76
  moving between, 75
clustering
  defined, 10, 175, 200
  density estimation, 196–199
  elbow method, 185–186
  expectation maximization (EM), 176, 192–196
  model goodness indicators, 195–196
  R language, 394–395
  in unsupervised learning, 10
clustering algorithms
  agglomerative, 175, 186–191
  divisive, 175–186
cognitive bias, 354
collaborative filtering (CF), 7–8
COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), 348
comprehensibility (machine learning), 359, 363
conditional probability, 382
confidence. *See* accuracy
content filtering, 8–9
context, 327
contextual bandits, 316–317
continuous bag of words (CBOW), 298
continuous variable, 145, 367
convolution, 247–248, 260
convolution neural network (CNN)
  and MLP, 247
  overview, 247
  structure, 248
Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), 348
correlation
  defined, 45, 377
  example, 378
  Pearson's *r* correlation, 378
  Python, 44
counterfactual evaluation, 340–343, 345
counterfactual explanation, 360, 363
coverage (association rule), 128

data bias, 354
data mining, 330–331, 345
data types
  defined, 30–31
  Python, 30
dataframe, Python, 54–55
decision rule, 127
decision tree. *See also* classification
  association rule, 128–129
  classification rule, 45, 128–129
  decision rule, 127
  entropy, 124–125
  information gain, 126
  purpose of, 123–124
  steps to create, 127
deep learning. *See also* machine learning
  classifier building, 272–278
  defined, 264, 298
  embeddings, 278–282
  encoders and transformers, 283–297
  as next level of machine learning, 264–265
  popularity of, 297–298
  Python model, 266
  shortcomings, 298
  simple model, 267–272
  when to use, 265–266
deep networks, 264, 298
dendogram
  defined, 189, 200
  with maximum clustering, 187
  with minimum clustering, 188
dendrites, 228
density estimation
  defined, 10
  Meanshift technique, 196–197
dependent variable (linear regression), 82, 114, 367
differential calculus functions, 380
differential privacy, 362
dimensionality, 222
dimensionality reduction
  curse of, 206–207
  feature selection, 207–213
  linear discrimination analysis (LDA), 388–389, 218–221
  maximum likelihood estimation (MLE), 213
  overfitting, 207
  principal component analysis (PCA), 214–218
discrete variable, 9, 367

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-009-12330-3 — A Hands-On Introduction to Machine Learning
Chirag Shah
Index
More Information

disinformation, 363
distributed filesystem (Hadoop), 54
distributed learning, 362
divisive clustering
  with *k*-means, 179–182
  with *k*-modes, 182–185
  steps to create, 176

elbow method (clustering), 185–186
elif command (Python), 32
embeddings
  deep learning and, 278
  defined, 278, 298
  input, 285
  pretrained word embeddings, 279
  word embeddings, 278–279
encoders
  layers in, 286–287
  positional, 285
Enron scandal, 356
ensemble methods, 132, 138
entropy
  decision tree, 124–125
  defined, 138
  types of, 126–127
epsilon-greedy algorithm, 317
estimator choice modeling criteria, 27, 112
evaluation (machine learning)
  A/B testing, 337–340
  adversarial learning, 344
  counterfactual evaluation, 340–343
  goodness and, 331
  offline, 331–335
  user research, 335–337
expectation maximization (EM)
  definition, 192
  log likelihood, 192
  maximum likelihood estimation (MLE), 192
  tossing coin example, 192
  uses for, 192
explainability (machine learning), 359, 360, 363
exploding gradient (RNN), 256
exploration–exploitation dilemma, 315–316

Fairness, Accountability, Transparency (FAT) conference, 352
Fairness, Accountability, Transparency, Ethics (FATE), 352
fairness versus bias in machine learning, 355–356
false negative statement, 332
false positive rate (FPR), 166
false positive statement, 332
feature, defined, 114, 222
feature selection
  defined, 222
  importance of, 207
  univariate feature selection, 208–210
federated learning, 362
feed-forward neural network, 231
filesystem, 54
F-measure, 332
focus groups, 336–337
for loop (Python), 32
forget gate (LTSM), 257
forward propagation methods (neural networks), 236

generalizability (model), 337
Google biases, 19
Google Cloud Platform (GCP)
  background, 50
  Google Colab, 56–59
  Hadoop, 53–55
  new project creation, 50–51
  SSG key addition, 52–53
  virtual machine creation, 50–51
Google Colab
  accessibility of, 56
  getting started running with Python, 56–59
  installing Python packages, 37–38
  versus Microsoft Azure, 66
Google or Not misinformation, 349–351
gradient ascent, 147
gradient descent, 98–102, 113, 114
graph-based neural networks (GNN), 297

Hadoop
  as backbone of big-data operations, 53–54
  defined, 54
  modules, 54–55
Hadoop Common (Hadoop), 55
HD Insight cluster, 60–61
Hello World program (Java), 14
hidden layer, 235
histogram, 368–369
hyperparameters modeling criteria, 112
hyperplane, 159–160

if command (Python), 32
if–else command (Python), 32–33
incremental gradient descent. *See* stochastic gradient descent
independent variable (linear regression), 82, 114, 367
information gain
  decision tree, 125
  defined, 138
input embeddings, 285
input gate (LSTM), 257
integrated development environment (IDE)
  defined, 45
  Python language access, 27–28
interim gradient descent. *See* stochastic gradient descent
interpretability (machine learning), 359, 363
interval variable, 367
interviews, 336–337
inverse propensity score (IPS), 342, 345

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-009-12330-3 — A Hands-On Introduction to Machine Learning
Chirag Shah
Index
More Information

Java language
  Hello World program, 14
  versus Python, 14
Jupyter Notebook
  installing and configuring, 383–384
  installing Python packages, 37–38

kernels
  decision boundaries, 166
  in SVM, 164
k-means
  divisive clustering, 179–182
  elbow method (clustering), 185–186
k-modes, 182–185
k-nearest neighbors (kNN)
  anomaly detection, 164–165
  easiest to understand and implement technique, 120
  example (trained model), 119–120
  major steps, 119
  reasons for, 119
  shortcomings, 120

learning
  limitations with computer association, 3
  positive change in system or person, 4, 20
learning rate, 30, 114
learning to rank (LTR), 329
likelihood of model, 146–147
linear discrimination analysis (LDA), 219
linear model, defined, 113
linear regression. *See also* regression
  defined, 45
  ordinary least squares (OLS), 84–85
  overfitting, 90
  overview, 90
  process of, 82
  R language, 390–391
  relationship between variables, 83
  statistical bias, 90
  terminology, 82–83
  using Python, 86–89
  variance, 90
linearity modeling criteria, 111
log likelihood, 192, 196, 200
logical operators (Python), 30
logistic regression
  categorical variable, 145
  continuous variable, 145
  defined, 147
  gradient ascent, 147
  likelihood of model, 146–147
  receiver operating curve (ROC), 151
long short-term memory (LSTM)
  defined, 257
  gates, 257

machine learned ranking (MLR), 329
machine learning. *See also* deep learning
  algorithms adjusting own parameters, 4–5
  defined, 3, 20, 113
  differences with artificial intelligence (AI), 10, 30
  learn from experience, 4
  neural network timing, 256
  nomenclature differences with statistics, 9
  regression at scale, 95
  traditional versus machine learning, 5
machine learning careers
  finance, 401
  healthcare, 401–402
  overview, 400–401
  retail, 402–403
  tech, 403
machine learning design and evaluation
  complexity of, 326
  evaluation, 331–344
  learning to rank (LTR) or machine learned ranking (MLR), 329
  recommender system, 327–328
machine learning example
  collaborative filtering (CF), 7–8
  content filtering, 8–9
  optical character recognition (OCR), 6
  WAYMO car, 5–6
machine learning issues
  fairness, 19
  flawed data, 18
  racial disparity, 18
  user exposure, 19
machine learning modeling criteria
  accuracy, 111
  estimator choice, 27, 112
  features numbers compared to data points, 112
  hyperparameters, 112
  linearity, 111
  training/testing, 111
  training time, 111
machine learning on the job
  certification for cloud platform, 74
  onboarding phase, 74
machine learning problems
  accountability, 356–358
  Beauty AI, 348–349
  bias, 353–356
  Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), 348
  differential privacy, 362
  federated learning, 362
  Google or Not misinformation, 349–351
  responsible AI in academia, industry, and regulations, 352–353
  transparency, 358–360
machine learning, specialized requirements
  skills, 11–13
  tools, 13–14

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-009-12330-3 — A Hands-On Introduction to Machine Learning
Chirag Shah
Index
More Information

machine learning taxonomy
  reinforcement, 10
  summary, 9, 20
  supervised, 9
  unsupervised, 10
MapReduce (Hadoop), 55
Markov process (reinforcement learning), 306
maximum likelihood estimation (MLE)
  defined, 200
  dimensionality reduction, 213
  expectation maximization (EM), 192
maximum marginal hyperplane (MMH), 160–161
McCulloch–Pitts model (MP), 229–230
mean (statistics), 372–373
Meanshift technique, 196–197
median (statistics), 373
Microsoft Azure
  account set-up, 59–60
  versus Amazon Web Services (AWS) Cloud9, 66
  cluster details overview, 62
  versus Google Colab, 66
  HDInsight cluster and, 60–61
  interface, 60
  as Linux virtual machine, 62
  machine learning example, 63–65
  storage set-up, 61–62
misinformation, 363
mode (statistics), 373–374
model, defined, 113
multi-arm bandit problem
  background, 316
  solutions, 317–322
multi-headed attention, 286–287
multilayer perceptron (MLP) model
  forward propagation methods, 236
  forward propagation methods example, 237–238
  hidden layer, 235
  layers in, 236
  overview of, 236
multinomial logistic regression, 152. *See also* softmax regression
multiple linear regression, 91
MySQL
  Python and, 396–398
  R language and, 398–399

naïve Bayes
  assumption, 154, 165–166
  defined, 154
  process of, 156
  steps to create, 155
neural networks
  backward propagation methods, 236–237
  basics of, 228
  classification example, 240–242
  construction in Python, 238
  design, 242–243
  forward propagation methods, 236
  graph-based (GNN), 297
  and human brain modeling, 227–228
  machine learning timing, 256
  regression example, 240
neural networks architecture
  autoencoder, 243–247
  convolution neural network (CNN), 247–251
  feed-forward network, 231
  long short-term memory (LSTM), 256–259
  multilayer perceptron (MLP), 234–235
  perceptron, 231–232
  recurrent neural network (RNN), 251–256
  single-layer perceptron (SLP), 232–235
neuron, defined, 260
neuron network architectures
  McCulloch–Pitts model (MP), 229–230
  single neuron, 230
  small, 229–230
neuron physiology
  axon, 228
  dendrites, 228
  and human processing power, 231
  structure, 229
  synapse, 228
nominal variable, 367
normal distribution, 370–372
NumPy package (Python), 39

offline evaluation metrics, 332–333
one-armed bandits, 315
optical character recognition (OCR), 6
ordinal variable, 367
ordinary least squares (OLS), 84–85
outcome. *See* dependent variable
out-of-bag error (random forest), 133
out-of-bag samples (random forest), 133
output gate (LSTM), 257
overfitting
  decision tree problem, 132
  dimensionality reduction, 207
  linear regression, 90
overfitting, data, 137

Pandas dataframe, 53–54
Pandas packages, 36
parallel distributed processing (PDP), 227–228
parameter, 114
Pearson's *r* correlation, 378
perceptron model
  defined, 231
  importance of, 231–232
  model representation, 232
  multilayer perceptron (MLP), 234–235
  single-layer perceptron (SLP), 232–235

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-009-12330-3 — A Hands-On Introduction to Machine Learning
Chirag Shah
Index
More Information

policies (reinforcement learning), 306
positional encoders, 285
positive predictive value. *See* precision metrics
precision metrics
  defined, 332
  example, 332–333
  F-measure, 332
  recall, 332
  receiver operating characteristic (ROC) curve, 334–335
predictor. *See* independent variable
predictor variable, 367
pretrained word embeddings, 279
principal component analysis (PCA)
  algorithm, 214–215
  defined, 214, 214
  process of, 214
  singular value decomposition (SVD), 215
probability formulas
  Bayes' theorem, 382
  conditional probability, 382
  event probability, 382
propensity, 342, 345
PuTTY, AWS connect, 68–69
PuTTYgen, 52
Python language
  basic examples, 29–31
  basic operation, 31–32
  control structures, 32–33
  deep learning model, 266
  and Google Colab, 56–59
  versus Java language, 14
  Jupyter Notebook, 383–384
  linear regression, 86–89
  neural network construction, 238
  overview, 26
  usefulness in a work environment, 44
Python language access
  Anaconda, 28, 383–384
  download/install, 26–27
  MySQL and, 396–398
  platforms, 26
  Spyder, 385, 29
  via console, 27
  via integrated development environment (IDE), 27–28
Python language functions
  form of, 35
  interactive, 36
  write own, 34–35
Python language packages
  matplotlib.pyplot, 40
  NumPy, 34, 39
  Pandas, 36
  using Google Colab, 37–38
  using Jupyter Notebook, 37–38
  using Spyder, 37
Python statistics
  bar graph, 41
  correlation, 44
  data distribution, 39
  histogram, 40–41
  importing data, 42–43
  max, 39
  mean, 39
  median, 40
  min, 39
  plotting data, 43–44
  standard deviation, 40
  variance, 40

Q-learning
  defined, 311
  goal of, 311
  process of, 311
Q-learning set-up
  Q matrix, 312
  reward matrix, 311
  states, 311

R language
  classification, 388–389, 392–393
  clustering, 394–395
  installing and configuring, 386
  linear regression, 390–391
  MySQL and, 398–399
  regression, 389–391
  unsupervised learning, 392–395
random forest
  advantages, 132, 134–135
  bootstrap sampling technique, 133
  versus decision trees, 50, 133–136
  reasons for, 132
  steps to create, 132
ratio variable, 367
recall, 332
receiver operating curve (ROC)
  Akaike Information Criterion (AIC), 334–335
  Bayesian Information Criterion (BIC), 335
  defined, 334
  generation, 151
  graph, 334
recurrent neural network (RNN)
  architecture, 20, 252
  defined, 252, 260
  problems with, 256–257
regression. *See also* linear regression
  defined, 81
  entry point to machine learning, 113
  multiple linear, 91–95
  neural networks example, 240
  R language, 389–391
  Ridge and Lasso, 96–98

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-009-12330-3 — A Hands-On Introduction to Machine Learning
Chirag Shah
Index
More Information

at scale, 95
in supervised learning, 9
reinforcement learning
ASIMO robot, 306
bandits, 315–322
as branch of machine learning, 10, 303
conceptual model, 304–306
cumulative award maximization, 304
defined, 20, 323
Markov process, 306
policies, 306
Q-learning, 311–315
tic tac toe, 306
training limit, 306
trial-and-error steps, 303
Ridge and Lasso regression
batch gradient descent, 20, 102
gradient descent, 98–102
reason to use, 98
stochastic gradient descent, 102

scatterplot (Python), 43–44
self-attention (transformers), 283, 284
Simpson's paradox, 355
single-layer perceptron (SLP) model
defined, 232
example for classification, 234–235
model representation, 233
overview, 232–233
singular value decomposition (SVD), 215
singular values (SVD), 215
skills for machine learning competence
computational thinking, 15–18
computer and data literacy, 12
critical thinking, 12–13
machine learning algorithms, 12
math, 12
to obtain job, 13
programming, 12
skip-gram models, 298
Spyder
installing and configuring, 385
installing Python packages, 37
plotting data, 43–44
Python language access, 29
SSH (secure client), 76
statistical bias, linear regression, 90
statistical distributions
central tendency, 372–374
dispersion, 374–377
histogram, 368–369
normal distribution, 370–372
statistical variables
continuous variable, 367
dependent variable (linear regression), 367
discrete variable, 367
independent variable (linear regression), 367
interval variable, 367
nominal variable, 367
ordinal variable, 367
ratio variable, 367
statistics
log likelihood, 192
maximum likelihood estimation (MLE), 192
nomenclature differences with machine learning, 9
receiver operating curve (ROC), 151
stochastic gradient descent, 102, 114
supervised learning
ANN, 228
anomaly detection, 164–165
classification, 118–138
decision tree, 123–132
defined, 9, 20, 113, 137, 166, 323
discrete, 9
know truth first in, 9–10
predictions about future, 118
R language, 392–393
random forest, 132–137
regression, 9, 81–113
support vector machine (SVM)
defined, 159
hyperplane, 159–160
kernels, 164
maximum marginal hyperplane (MMH), 160–161
theory, 161–162
types of data, 159
surveys, 336
synapse, 228, 26, 260
systemic bias, 353–354

test set, 137, 146
Thompson sampling, 317
tic tac toe (reinforcement learning), 306
tools for machine learning competence
Python language, 14
R language, 386–395
training limit (reinforcement learning), 306
training time modeling criteria, 111
training–validation–test data, 137, 146, 166
transformers
Bidirectional Encoder Representations from Transformers (BERT), 287–297
defined, 283
input embeddings, 285
self-attention, 283, 284
transparency (machine learning), 358–360, 363
true negative statement, 332
true positive rate (TPR), 166
true positive statement, 332

understandability, 363
understandability (machine learning), 359
univariate feature selection, 208–210

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-009-12330-3 — A Hands-On Introduction to Machine Learning
Chirag Shah
Index
More Information

unsupervised learning
  autoencoder network, 244–246
  clustering, 175–199
  clustering in, 10
  defined, 200, 323
  density estimation, 10
  dimensionality reduction, 206–221
  R language, 394–395
upper confidence bound, 317
user research (machine learning evaluation)
  interviews, 336–337
  surveys, 336
  timing for, 335
  user studies, 337
user studies, 337

validation set, 137
vanishing gradient (RNN), 256
variance (linear regression), 90

WAYMO car, 5–6
while loop (Python), 32, 33
word embeddings, 278–279

YARN (Hadoop), 55

© in this web service Cambridge University Press & Assessment www.cambridge.org