INDEX

A

activation function, 171–172
Adadelta, 313, 326, 344, 350
Adagrad, 313, 326
Adam, 313, 326
Apache MXNet, 2
area under the ROC curve (AUC), 273–274, 354
array processing, 2
artificial intelligence (AI), xxviii
Arxiv, 414
Az, 273

B

backpropagation, 174, 190, 200, 203
bagging, 123
batch size, 231
batch training, 195
Bayes, Thomas, 114
Bengio, Yoshua, 416
big-O notation, 155
bootstrap sample, 123
bounding box, 340
box plot, 78
  IQR, 80
  whiskers, 80
Brox, Thomas, 304
Burges, Christopher, 126

C

Caffe, 2
Caffe2, 2
capsule networks, 301
Cartesian plane, 4
case study, 378
  classical models, 385
  classifier
    audio features, 385
    neural network, 388
    spectrograms, 398
  CNN
    batch normalization, 402
    ensembles, 404
    initialization, 402
    one-dimensional, 389
    regularization, 402
    two-dimensional, 398
    voting, 405
  dataset, 378
    augmentation, 379
    ESC-10, 378
    preprocessing, 383
    spectrogram, 394–395
  `sox`, 396
categorical value, 54
centroid, 109
CIFAR-10, 90, 343–344
  binary versus multiclass, 357
  building models, 345
  one-versus-rest, 357
classes, 51
classification, 51
clustering algorithms, 413
CNN, 6, 283
  activations, 306
  advanced models, 411
  AlexNet, 412
  anatomy of, 288
  architecture, 288–289
  bounding box, 340
  building, 310
  constructing in Keras, 312
  convolution, 284
    exact, 287
    filter, 293
    valid, 287
    zero-padding, 287
  convolutional layer
    effect, 295
    initialization, 299
    operation, 292
    stacking, 298

CNN, *continued*
data flow, 291
DenseNet, 412
effective receptive field, 298–299
end-to-end training, 307
evaluating, 314
experiments, 319
architecture, 319
epochs, 323
minibatches, 323
optimizers, 326
training set size, 323
feature maps, 290
filter, 293
fully convolutional network, 303, 309, 328, 331
building and training, 328
graymap, 337
heatmap, 336
history, 284
Inception, 412
kernel, 284
using, 285
layer types, 289
layers
convolutional, 289, 292
dense, 289
dropout, 289
flatten, 289
fully connected, 290, 301
fully convolutional, 302
inner product, 290
inputs, 290
pooling, 289, 299
ReLU, 289
LeNet, 412
loss function, 313
max pooling, 300
motivation, 284
Neocognitron, 284
optimizer, 313
Adadelta, 313
Adagrad, 313
Adam, 313
RMSprop, 313
stochastic gradient descent (SGD), 313
optimizers, 326
outputs, 304
plotting
error, 317
loss, 317
pooling layer
maximum, 300
stride, 299
probability map, 336
receptive field, 298
ResNet, 412
scaling data, 311
semantic segmentation, 412
spatial invariance, 284
stride, 285
training, 314
U-Net, 340, 412
VGG, 412
YOLO, 340, 412
Cohen's kappa, 263
comparing models, 271
Compute Unified Device Architecture (CUDA), 9, 412
conference
Computer Vision and Pattern Recognition (CVPR), 415
GPU Technology Conference (GTC), 415
International Conference on Learning Representations (ICLR), 415
International Conference on Machine Learning (ICML), 415
Neural Information Processing Systems (NIPS), 415
NeurIPS, 415
conferences, 415
confusion matrix, 254, 255
multiclass, 276
contingency table, 255
convolution, 284
exact, 287
image processing, 287
valid, 287
zero-padding, 287
convolutional neural network
see CNN, 283
Cortes, Corinna, 124
Coursera, 414
Courville, Aaron, 416

curse of dimensionality, 55
curve fitting, 213

D

data augmentation, 92, 215
  approaches, 94
  images, 101
  rationale, 93
  using PCA, 97
dataset, 7, 57
  k-fold validation, 74
  augmentation, 92
    approaches, 94
    images, 101
    PCA, 97
    rationale, 93
  bagging, 123
  balanced, 64
  breast cancer dataset, 86
  categorical value, 54
  cautionary tales, 80
  CIFAR-10 dataset, 90
  classes, 51
  confuser, 61–62
  curse of dimensionality, 55
  discrete value, 53
  ESC-10, 378
  feature selection, 55
  feature vectors, 53
  floating-point number, 53
  hard negative, 62
  interval value, 53
  irises dataset, 84
  label, 51
  mean centering, 64
  mislabeled data, 76
  missing features, 67
  MNIST dataset, 88
  normalizing, 65–66, 135
  one-hot encoding, 55
  ordinal value, 54
  outliers, 76, 80
  partitioning, 69–70, 74
  partitioning by class, 70
  pitch shifting, 379
  potential problems, 76
  preprocessing, 63, 83
  random sampling, 72
  sample, 4
  scaling, 63–64
  spectrogram, 394–395
  standardizing, 65–66, 135
  summarizing, 76
  test data, 68
  time shifting, 379
  training data, 68
  validation data, 68
Decision Tree, 164
deep learning, xxviii
Deep Learning Book (Goodfellow et al.), 416
derivative, 192
  local minimum, 193
  minimum, 192
  partial, 201
  tangent line, 192
descriptive statistics, 7
discriminative network, 413
domain adaptation, 60, 331, 413
dot product, 5
dropout, 217

E

effective receptive field, 298–299
embedding, 361
ensemble, 123
epoch, 195
Euclidean distance, 110
experiments
  breast cancer, 135
  CIFAR-10
    analyzing models, 348
    animal or vehicle, 352
    binary versus multiclass, 357
    building models, 345
    fine-tuning, 367
    transfer learning, 361
  irises, 129, 182
    classical models, 130
  MNIST, 150
    activation function, 223
    activation results, 227
    architecture, 223
    architecture results, 227
    base learning rate, 235
    batch size, 231
    classical models, 150
    code, 223

experiments, MNIST, *continued*
initialization, 243
L2 regularization, 239
momentum, 242
neural networks, 222
scrambled, 161, 247
training set size, 238
MNIST CNN
basic experiments, 319
building models, 312
dataset, 310
epochs, 323
fully convolutional, 328
minibatches, 323
optimizers, 326
scrambled, 340
training set size, 323
extrapolation, 58

F

F1 score, 263
false negative rate (FNR), 259
false positive rate (FPR), 259
feature, 52
types, 53
feature selection, 55
feature vectors, 52–53
fine-tuning, 367
example, 368
rationale, 371
Fischer, Philipp, 304
Fisher, R.A., 84
floating-point number, 53, 85, 382
Fourier transform, 394
power spectrum, 394
Frank, Eibe, 157
fully connected layers, 301
fully convolutional layer, 302
fully convolutional network, 328
building and training, 328
function
convex, 196
nonconvex, 196

G

Gaussian distribution, 8
generative adversarial network
CycleGAN, 413
generative adversarial network (GAN), 413
generative network, 413
Gini index, 122
GitHub, 414
Glorot, Xavier, 244
Goodfellow, Ian, 413, 416
GPU, 1
gradient, 190
slope, 191
gradient descent, 190
algorithm, 194
batch training, 195
epoch, 195
first-order, 197
learning rate, 193, 235
local minimum, 193
minibatch, 195–196
momentum, 199, 242
stochastic, 194
grand mean, 143
graph, 170
edges, 170
nodes, 170
graphics processing unit (GPU), 1, 9

H

Hadamard product, 207
Hadamard, Jacques, 207
HDF5, 3, 314
He, Kaiming, 212
heatmap, 336
Hinton, Geoffrey E., 200, 301, 415
hit rate, 257
hyperbolic tangent, 174
hyperparameters, 144
optimizing, 145
hyperplane, 125
hypothesis testing, 8

I

image processing, 287
informedness, 262
inner product, 5
interpolation, 58
interval value, 53

J

joint probability, 115

K

K-D-tree, 157
k-fold validation, 74, 140
k-means, 413
k-Nearest Neighbor (*k*-NN), 112–113
k-NN classifier, 112–113, 139, 145, 154
Kaggle, 415
Karpathy, Andrej, 414
Keras, 2
  documentation, 311, 412
kernel, 284
Kibriya, Ashraf M., 157
Kneusel, Ronald T. *Numbers and Computers*, 14
Krizhevsky, Alex, 217

L

L2 regularization, 216
label, 51
learning rate, 193–194, 198, 235
  schedule, 198
LeCun, Yann, 412–413
line
  intercept, 191
  slope, 191
linear algebra, 4
  matrix, 5
    multiplication, 6
    multiplication by vector, 5
  scalar, 5
  tensor, 6
  transpose, 5
  vector, 4
    column, 4
    inner product, 5
    multiplication, 5
    outer product, 5
    row, 4
linear function, 172
linear regression, 58
Linux, 3
  Ubuntu, 3
local minimum, 193
long short-term memory, 414
loss function, 190, 192, 208, 313
  absolute error, 209
  cross-entropy loss, 210, 313
  mean squared error (MSE), 209
LSTM, 414

M

Macintosh, 3
macOS, 3
`main` function, 99
Mann-Whitney U test, 8, 147
markedness, 262
matrix, 5, 31
  multiplication, 6, 181
  multiplication by vector, 5
  transpose, 185
matrix multiplication, 6
Matthes, Eric, *Python Crash Course*, 11
Matthews Correlation Coefficient (MCC), 264
  multiclass, 281
max pooling, 300
mean, 6
  standard error, 6
median, 7
metric, 251
metrics
  2×2 table, 255
  accuracy, 132, 252
  accuracy matrix, 277
  advanced metrics, 262
  area under the ROC curve (AUC), 273
  Az, 273
  CIFAR-10 example, 354
  Cohen's kappa, 263
  confusion matrix, 254–255
    multiclass, 276
  contingency table, 255
  derived metrics, 257
  F1 score, 263
  false negative (FN), 254–255
  false negative rate (FNR), 259
  false positive (FP), 254–255
  false positive rate (FPR), 259
  grand mean, 143
  hit rate, 257
  implementation, 264
  informedness, 262
  interpreting models, 260

metrics, *continued*
markedness, 262
Matthews Correlation Coefficient (MCC), 264
multiclass, 281
multiclass, 276
negative predictive value (NPV), 258
per class accuracy, 253
positive predictive value (PPV), 258
precision, 258
precision-recall curve, 275
recall, 257
receiver operating characteristics (ROC), 266
elements of, 269
generating, 273
score, 132
sensitivity, 257
specificity, 257
t-SNE, 307, 363
true negative (TN), 254
true negative rate (TNR), 257
true positive (TP), 254, 254
true positive rate (TPR), 257
weighted accuracy, 279
Youden's *J* statistic, 262
minibatch, 195–196, 231
minibatch training, 196
missing features, 67
model
capacity, 62
classical models, 108
computational requirements, 165
explainability, 166
small datasets, 165
summary, 162
vector inputs, 166
when to use, 165
comparing, 271
Decision Tree, 117–118
construction, 121
summary, 164
ensemble, 123
fine-tuning, 367
generative adversarial network (GAN), 413
Gini index, 122
hyperparameters, 144
optimizing, 145
long short-term memory, 414
LSTM, 414
Naïve Bayes, 113
Gaussian, 131
Multinomial, 131
summary, 163
Nearest Centroid, 108, 111
summary, 162
Nearest Neighbor, 5, 112
optimizing, 145
summary, 163
overfitting, 190, 213
pretrained, 368
Random Forest, 117, 122–123
optimizing, 146
summary, 164
recurrent neural network (RNN), 414
Support Vector Machine (SVM), 124
kernel, 127
margin, 124–125
optimizing, 126, 147
summary, 165
support vector, 126
template matching, 111
transfer learning, 361
momentum, 199, 242
multidimensional array, 30

N

Naïve Bayes, 115, 155, 163
Gaussian, 144, 155
Nearest Centroid, 162
Nearest Neighbor, 5, 163
negative predictive value (NPV), 258
neural network, 171
activation function, 171–172
hyperbolic tangent, 174
identity, 178
ReLU, 174
sigmoid, 174
anatomy, 170
ANN, 172
architecture, 176
backpropagation, 190, 200
backward pass, 200
bias, 172
essence, 172
feedback, 171

forward pass, 200
fully connected feedforward, 169, 170
function approximation, 170
hidden layer, 170
implementation, 182–183
multi-layer perceptron (MLP), 172
neuron, 170–171
output layer, 172, 178
regularization, 190, 213, 215
dropout, 217
L2, 216
weight decay, 216
representation, 180
rules of thumb, 177
softmax, 178
traditional, 169
training, 219
batch, 195
epoch, 195
gradient descent, 190
loss function, 190, 192, 208
minibatch, 195, 196
momentum, 199
overview, 190
stochastic gradient descent (SGD), 194
stopping, 197
weight update, 193
weight initialization, 190, 211, 243
weights, 172
Ng, Andrew, 414
nominal value, 54
nonlinear function, 172
transcendental, 173
trigonometric, 173
nonparametric test, 8
normal distribution, 7, 8
normalizing, 65–66, 135
Numbers and Computers (Kneusel), 14
NumPy, 2
`arange`, 39–40
`argsort`, 72
array, 33
array indexing, 37
arrays versus lists, 30
`astype`, 37
basic arrays, 33
broadcasting, 42
contiguous memory, 30
convert array to image, 104
convert image to array, 104
`copy`, 37
data types, 33, 34
documentation, 29
`dot`, 44
`dtype`, 33
ellipsis, 41, 42
histogram, 356
images, 48
channels, 48
indexing and memory, 30
input and output, 45
`load`, 46
`loadtxt`, 46
normally distributed random numbers, 48
`ones`, 36–37
operators, 42
performance, 31
pseudorandom seed, 138
random
`seed`, 48
`random`, 48
rationale, 29
`reshape`, 40
`save`, 46
`savetxt`, 46
`savez`, 46
`savez_compressed`, 47
`shape`, 35
slicing, 39
short cuts, 40
subscript, 38
uniform random numbers, 48
`where`, 70
`zeros`, 36
NVIDIA, 9, 412

O

one-hot encoding, 55
online resources, 414
operating environment, 1
ordinal value, 54
outer product, 5
outliers, 80
overfitting, 94, 198, 213

P

p-value, 8
parent distribution, 7, 60, 252
partial derivative, 201
Pasteur, Louis, 167
Piczal, Karol J., 378
Pillow, 3
  `PIL`, 48
  `convert`, 50
  `fromarray`, 49
  `open`, 49
  `save`, 49
  `show`, 49
  documentation, 50
Plato, 7, 60
pooling, 299
positive predictive value (PPV), 258
power spectrum, 394
Powers, David Martin, 262
precision, 258
precision-recall curve, 275
preprocessing, 83
primary visual cortex, 297
principal component analysis (PCA), 95, 153, 158
  MNIST, 158
prior class probability, 60, 253
probability, 6
  Bayes' theorem, 114
  distribution, 7
  Gaussian distribution, 8
  joint probability, 115
  likelihood, 114
  normal distribution, 7, 8, 116
  notation
    uniform distribution, 8
  parent distribution, 7, 60
  posterior probability, 114
  prior class probability, 60
  prior probability, 114
  uniform distribution, 7
probability distribution, 7
probability map, 336
pseudorandom sequence, 140
Python
  blocks, 12
  bottom-tested loop, 22
  `break`, 22–23
  `continue`, 22–23
control structures, 19
  `for`, 20
  `if`, 12, 19
  `try`, 24
  `while`, 22
  `with`, 23
data structures, 13
debugging, 24
dictionaries, 13, 18
documentation, 27
dynamic typing, 14
editor conventions, 13
`enumerate`, 20
essence, 11
`except`, 24
exceptions, 24
exiting, 12
`False`, 16
floating-point numbers, 13
`for` loops, 20
function definition, 24
  defaults, 25
if statement, 19
intentation, 12
interactive mode, 12
`librosa`, 380
list comprehensions, 21
list operations, 15
lists, 13, 15
  copying, 16
lists in memory, 30
module, 26
  `copy`, 17
  `deepcopy`, 17
  `time`, 26, 32
namespaces, 26
None, 16
`pass`, 13
`pickle`, 45, 184
primitive data types, 15
`range`, 20, 39
statements, 12
strings, 13, 14
  quotes, 15
top-tested loop, 22
`True`, 16
`try`, 24
tuple, 17
variables, 13–14

camel-case, 14
naming, 14
while loop, 22
whitespace, 12
with statement, 23
Python Crash Course (Matthes), 11
PyTorch, 2

R

Raghu, Maithra, 375
Random Forest, 122–123, 164
Random Numbers and Computers (Kneusel), 48
recall, 257
receiver operating characteristics (ROC), 266, 269–275
elements of, 269
receptive field, 298
rectified linear unit (ReLU), 174
recurrent neural networks (RNNs), 414
recursion, 120
Reddit Machine Learning, 414
regularization, 94, 190, 213, 215
dropout, 217
L2, 216, 239
weight decay, 216, 239
reinforcement learning, xxviii, 412
AlphaGo, 413
AlphaGo Zero, 413
Atari 2600, 413
RMSprop, 313, 326
Ronneberger, Olaf, 304
Rumelhart, David E., 200

S

sample, 4
scalar, 5
scikit-learn, 2
DecisionTreeClassifier, 130
documentation, 2
GaussianNB, 130
KNeighborsClassifier, 130
make_classification, 70
MLPClassifier, 172, 186, 222
MultinomialNB, 130
NearestCentroid, 130
RandomForestClassifier, 130
SVC, 130
SciPy, 380
wavfile, 380
semantic segmentation, 412
sensitivity, 257
sigmoid, 174
softmax, 178
specificity, 257
spectrogram, 395
standard deviation, 7–8, 32, 65, 211
standard error, 6
standardizing, 65–66, 135
statistical test, 8
statistically significant, 9
statistics, 6
bootstrap sample, 123
descriptive, 6–7
dot product, 44
grand mean, 143
hypothesis testing, 8
Mann-Whitney U test, 8
mean, 6
standard error, 6
median, 7
nonparametric test, 8
p-value, 8
parametric test, 8
quartile, 79
standard deviation, 7, 8, 33, 65, 117
standard error, 6, 32, 77
statistically independent, 115
statistically significant, 9
t-test, 8
variance, 7
stochastic gradient descent (SGD), 194, 224, 313, 326, 344
supervised learning, xxviii, 57, 412
Support Vector Machine (SVM), 165
kernel, 127
margin, 124–125
one-versus-one, 156
one-versus-rest, 156
optimizing, 126
support vector, 126

T

t-SNE, 307
t-test, 8
tangent line, 192

tensor, 6
TensorFlow, 2
  documentation, 2
test data, 68
toolkit
  Apache MXNet, 2
  Caffe, 2
  Caffe2, 2
  HDF5, 3
  installation, 3
  Keras, 2
  NumPy, 2
  Pillow, 3
  PyTorch, 2
  scikit-learn, 2
  TensorFlow, 2
  versions, 3
training data, 68
training set, xxviii
transcendental functions, 173
transfer learning, 361
  embedding, 361
trigonometric functions, 173
true negative rate (TNR), 257
true positive rate (TPR), 257

U

U-Net, 340
Ubuntu, 1
UCI Machine Learning Repository, 84
uniform distribution, 7
uniformly distributed, 211
unsupervised learning, xxviii, 412

V

validation data, 68
Vapnik, Vladimir, 124
variance, 7
vector, 4, 31
  column vector, 4
  dot product, 5
  inner product, 5
  multiplication, 5
  outer product, 5
  row vector, 4

W

weight decay, 216
weight initialization, 190, 211, 243
  Glorot, 212
  He, 212
  Xavier, 212
weight update, 193–194
weighted accuracy, 279
weights, 172
Williams, Ronald J., 200
Windows, 3

Y

YOLO, 340
Youden's *J* statistic, 262
YouTube, 415

Z

zero-padding, 287