

INDEX

Symbols

k-fold validation, 75, 140
k-means, 401

A

activation function, 170
Adadelta, 307, 320, 338, 344
Adagrad, 307, 320
Adam, 307, 320
area under the ROC curve, 267
array processing, 2
Arxiv, 402
AUC, 267
Az, 267

B

backpropagation, 172, 186, 195, 196
bagging, 122
batch size, 226
batch training, 190, 191
Bayes, Thomas, 113
Bengio, Yoshua, 404
big-O notation, 155
bootstrap sample, 122
bounding box, 333
box plot, 79
 IQR, 80
 whiskers, 80

C

Caffe, 2
Caffe2, 2
capsule networks, 294
Cartesian plane, 4
case study, 368
 classical models, 375
 classifier
 audio features, 375
 neural network, 378
 spectrograms, 387

CNN

 batch normalization, 389
 ensembles, 392
 initialization, 389
 one-dimensional, 379
 regularization, 389
 two-dimensional, 387
 voting, 393
dataset, 368
 augmentation, 369
 ESC-10, 368
 preprocessing, 373
 spectrogram, 383, 384
 sox, 385
categorical value, 56
centroid, 109
CIFAR-10, 90, 337, 338
 binary versus multiclass, 348
 building models, 339
 one-versus-rest, 349
classes, 53
classification, 53
clustering algorithms, 401
CNN, 277
 activations, 299
 advanced models, 399
 AlexNet, 400
 anatomy of, 282
 architecture, 282
 bounding box, 333
 building, 304
 constructing in Keras, 306
 convolution, 278
 exact, 280
 filter, 286
 valid, 280
 zero-padding, 280
 convolutional layer
 effect, 288
 initialization, 292
 operation, 286
 stacking, 291
data flow, 285

- DenseNet, 400
- effective receptive field, 291
- end-to-end training, 300
- evaluating, 307
- experiments, 313
 - architecture, 313
 - epochs, 317
 - minibatches, 317
 - optimizers, 320
 - training set size, 317
- feature maps, 284
- filter, 286
- fully convolutional network, 296, 303, 321, 323
 - building and training, 321
- graymap, 330
- heatmap, 329
- history, 278
- Inception, 400
- kernel, 278
 - using, 279
- layer types, 283
- layers
 - convolutional, 283, 286
 - dense, 283
 - dropout, 283
 - flatten, 283
 - fully connected, 283, 294
 - fully convolutional, 295
 - inner product, 283
 - inputs, 284
 - pooling, 283, 292
 - ReLU, 283
- LeNet, 400
- loss function, 307
- max pooling, 293
- motivation, 278
- Neocognitron, 278
- optimizer, 307
 - Adadelta, 307
 - Adagrad, 307
 - Adam, 307
 - RMSprop, 307
 - stochastic gradient descent, 307
- optimizers, 320
- outputs, 297
- plotting
 - error, 310
 - loss, 310
- pooling layer
 - maximum, 293
 - stride, 292
- probability map, 329
- receptive field, 291
- ResNet, 400
- scaling data, 305
- semantic segmentation, 400
- spatial invariance, 278
- stride, 279
- training, 307
- U-Net, 333, 400
- VGG, 400
- YOLO, 333, 400
- Cohen's kappa, 257
- comparing models, 265
- Compute Unified Device Architecture, 9
- conference, 403
 - CVPR, 403
 - GTC, 403
 - ICLR, 403
 - ICML, 403
 - NeurIPS, 403
 - NIPS, 403
- confusion matrix, 248, 249
 - multiclass, 270
- contingency table, 249
- convolution, 278
 - exact, 280
 - image processing, 281
 - valid, 280
 - zero-padding, 280
- convolutional neural network, 6
 - see CNN, 277
- Coursera, 402
- Courville, Aaron, 404
- CUDA, 9, 400
- curse of dimensionality, 57
- curve fitting, 207

D

- data augmentation, 92, 210
 - approaches, 94
 - images, 101
 - rationale, 93
 - using PCA, 97
- dataset, 7, 58
 - k -fold validation, 75

- augmentation, 92
 - approaches, 94
 - images, 101
 - PCA, 97
 - rationale, 93
- bagging, 122
- balanced, 64
- breast cancer dataset, 86
- categorical value, 56
- cautionary tales, 81
- CIFAR-10 dataset, 90
- classes, 53
- confuser, 62, 63
- curse of dimensionality, 57
- discrete value, 55
- ESC-10, 368
- feature selection, 57
- feature vectors, 55
- floating-point number, 55
- hard negative, 63
- interval value, 55
- irises dataset, 84
- label, 54
- mean centering, 65
- misabeled data, 77
- missing features, 68
- MNIST dataset, 88
- normalizing, 66, 67, 135
- one-hot encoding, 57
- ordinal value, 56
- outliers, 77, 80
- partitioning, 70, 74
- partitioning by class, 71
- pitch shifting, 369
- potential problems, 76
- preprocessing, 64, 83
- random sampling, 73
- sample, 4
- scaling, 64, 65
- spectrogram, 383, 384
- standardizing, 66, 67, 135
- summarizing, 77
- test data, 69
- time shifting, 369
- training data, 69
- validation data, 69
- decision tree, 163
- Deep Learning Book, 404
- Deep Learning Book*, 404

- derivative, 188
 - local minimum, 189
 - minimum, 188
 - partial, 196
 - tangent line, 188
- descriptive statistics, 7
- discriminative network, 401
- domain adaptation, 61, 324, 401
- dot product, 5
- dropout, 212

E

- effective receptive field, 291
- embedding, 353
- ensemble, 122
- epoch, 191
- Euclidean distance, 110
- experiments
 - breast cancer, 135
 - CIFAR-10
 - analyzing models, 342
 - animal or vehicle, 344
 - binary versus multiclass, 348
 - building models, 339
 - fine-tuning, 358
 - transfer learning, 352
 - irises, 130, 179
 - classical models, 130
 - MNIST, 150
 - activation function, 218
 - activation results, 222
 - architecture, 218
 - architecture results, 222
 - base learning rate, 229
 - batch size, 226
 - classical models, 150
 - code, 219
 - initialization, 237
 - L2 regularization, 233
 - momentum, 236
 - neural networks, 218
 - scrambled, 160, 241
 - training set size, 232
 - MNIST CNN
 - basic experiments, 313
 - building models, 306
 - dataset, 304
 - epochs, 317
 - fully convolutional, 321

- minibatches, 317
- optimizers, 320
- scrambled, 333
- training set size, 317

extrapolation, 59

F

F1 score, 256

false negative rate, 253

false positive rate, 253

feature, 54

- types, 55

feature selection, 57

feature vectors, 54, 55

fine-tuning, 358

- example, 359
- rationale, 362

floating-point number, 55

FNR, 253

Fourier transform, 383

- power spectrum, 383

FPR, 253

fully connected layers, 294

fully convolutional layer, 295

fully convolutional network, 321

- building and training, 321

function

- convex, 192
- nonconvex, 192

G

GAN, 401

Gaussian distribution, 8

generative adversarial network, 401

- CycleGAN, 401

generative network, 401

Gini index, 121

Github, 402

Glorot, Xavier, 238

Goodfellow, Ian, 401, 404

GPU, 1, 8

gradient, 186

- slope, 187

gradient descent, 186

- algorithm, 190
- batch training, 190, 191
- epoch, 191
- first-order, 193

- learning rate, 189, 230
- local minimum, 189
- minibatch, 190, 192
- momentum, 195, 236
- stochastic, 190

grand mean, 143

graph, 168

- edges, 168
- nodes, 168

graphics processing unit, 1, 8

H

Hadamard product, 201

HDF5, 3, 308

heatmap, 329

Hinton, Geoffrey, 294, 403

hit rate, 251

hyperbolic tangent, 172

hyperparameters, 144

- optimizing, 144

hyperplane, 124

hypothesis testing, 8

I

image processing, 281

informedness, 255

inner product, 5

interpolation, 59

interval value, 55

J

joint probability, 115

K

K-D-tree, 157

Kaggle, 403

Karpathy, Andrej, 402

Keras, 2

- documentation, 305, 400

kernel, 278

Kneusel, Ronald T. *Numbers and Computers*, 14

L

L2 regularization, 210

label, 54

learning rate, 189, 190, 194, 230

- schedule, 194
- LeCun, Yann, 400, 401
- line
 - intercept, 187
 - slope, 187
- 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, 171
- linear regression, 59
- Linux, 3
 - Ubuntu, 3
- local minimum, 189
- long short-term memory, 402
- loss function, 186, 188, 203, 307
 - absolute error, 203
 - cross-entropy loss, 204, 307
 - mean squared error (MSE), 203
- LSTM, 402

M

- Macintosh, 3
- macOS, 3
- Mann-Whitney U test, 8, 146
- markedness, 255
- matrix, 5, 31
 - multiplication, 6, 178
 - multiplication by vector, 5
 - transpose, 181
- matrix multiplication, 6
- Matthes, Eric, *Python Crash Course*, 11
- Matthews Correlation Coefficient, 257
 - multiclass, 274
- max pooling, 293
- MCC, 257
 - multiclass, 274
- mean, 6
 - standard error, 6
- median, 7

- metric, 245
- metrics
 - 2x2 table, 249
 - accuracy, 132, 246
 - accuracy matrix, 271
 - advanced metrics, 255
 - area under the ROC curve, 267
 - AUC, 267
 - Az, 267
 - CIFAR-10 example, 346
 - Cohen's kappa, 257
 - confusion matrix, 248, 249
 - multiclass, 270
 - contingency table, 249
 - derived metrics, 250
 - F1 score, 256
 - false negative, 248
 - false negative rate, 253
 - false positive, 248
 - false positive rate, 253
 - FN, 248
 - FNR, 253
 - FP, 248
 - FPR, 253
 - grand mean, 143
 - hit rate, 251
 - implementation, 258
 - informedness, 255
 - interpreting models, 253
 - markedness, 255
 - Matthews Correlation Coefficient, 257
 - multiclass, 274
 - MCC, 257
 - multiclass, 274
 - multiclass, 269
 - negative predictive value, 252
 - NPV, 252
 - per class accuracy, 247
 - positive predictive value, 252
 - PPV, 252
 - precision, 252
 - precision-recall curve, 269
 - recall, 251
 - receiver operating characteristics, 259
 - ROC, 259
 - elements of, 262, 263
 - generating, 267

- score, 132
- sensitivity, 251
- specificity, 251
- t-SNE, 300, 355
- TN, 248
- TNR, 251
- TP, 248
- TPR, 251
- true negative, 248
- true negative rate, 251
- true positive, 248
- true positive rate, 251
- weighted accuracy, 273
- Youden's J statistic, 255
- minibatch, 190, 192, 226
- minibatch training, 192
- missing features, 68
- model
 - capacity, 63
 - classical models, 108
 - computational requirements, 165
 - explainability, 165
 - small datasets, 164
 - summary, 161
 - vector inputs, 165
 - when to use, 164
 - comparing, 265
 - decision tree, 117, 118
 - construction, 120
 - summary, 163
 - ensemble, 122
 - fine-tuning, 358
 - GAN, 401
 - generative adversarial network, 401
 - Gini index, 121
 - hyperparameters, 144
 - optimizing, 144
 - long short-term memory, 402
 - LSTM, 402
 - naïve Bayes, 113
 - Gaussian, 131
 - Multinomial, 131
 - summary, 162
 - nearest centroid, 108, 111
 - summary, 161
 - nearest neighbor, 4, 112
 - optimizing, 144
 - summary, 162
 - overfitting, 186, 207
 - pretrained, 359
 - random forest, 117, 122
 - optimizing, 145
 - summary, 163
 - recurrent neural network, 402
 - Support Vector Machine (SVM), 123
 - kernel, 126
 - margin, 123, 125
 - optimizing, 126, 146
 - summary, 164
 - support vector, 125
 - template matching, 111
 - transfer learning, 352
- momentum, 195, 236
- MXnet, 2

N

- naïve Bayes, 162
- nearest centroid, 161
- nearest neighbor, 4, 162
- negative predictive value, 252
- neural network, 169
 - activation function, 170
 - hyperbolic tangent, 172
 - identity, 175
 - relu, 172
 - sigmoid, 172
- anatomy, 168
- ANN, 170
- architecture, 173
- backpropagation, 186, 195, 196
- backward pass, 196
- bias, 170
- essence, 170
- feedback, 169
- forward pass, 196
- fully-connected feedforward, 167, 168
- function approximation, 168
- hidden layer, 168
- implementation, 178, 180
- multi-layer perceptron (MLP), 170
- neuron, 168, 169
- output layer, 170, 175
- regularization, 186, 207, 210
 - dropout, 212
 - L2, 210
 - weight decay, 211
- representation, 177

- rules of thumb, 174
- softmax, 175
- traditional, 167
- training, 214
 - batch, 190, 191
 - epoch, 191
 - gradient descent, 186
 - loss function, 186, 188, 203
 - minibatch, 190, 192
 - momentum, 195
 - overview, 186
 - stochastic gradient descent, 190
 - stopping, 193
 - weight update, 189
- weight initialization, 186, 205, 237
- weights, 170
- Ng, Andrew, 402
- nominal value, 56
- nonlinear function, 170, 171
 - transcendental, 171
 - trigonometric, 171
- nonparametric test, 8
- normal distribution, 7, 8
- normalizing, 66, 67, 135
- NPV, 252
- Numbers and Computers* (Kneusel), 14
- NumPy, 2
 - arange, 39, 41
 - argsort, 72
 - array, 33
 - array indexing, 38
 - arrays versus lists, 30
 - astype, 37
 - basic arrays, 33
 - broadcasting, 43
 - contiguous memory, 30
 - convert array to image, 103
 - convert image to array, 104
 - copy, 38
 - data types, 34, 35
 - documentation, 29
 - dot, 44, 45
 - dtype, 34
 - ellipsis, 41, 43
 - histogram, 348
 - images, 49
 - channels, 49
 - indexing and memory, 30
 - input and output, 46
 - load, 47
 - loadtxt, 47
 - normally distributed random numbers, 48
 - ones, 36, 37
 - operators, 43
 - performance, 31
 - pseudorandom seed, 138
 - random
 - seed, 48
 - random, 48
 - rationale, 29
 - reshape, 41
 - save, 47
 - savetxt, 47
 - savez, 47
 - savez_compressed, 48
 - shape, 35, 36
 - slicing, 39
 - short cuts, 40
 - uniform random numbers, 48
 - where, 71
 - zeros, 36
- NVIDIA, 9, 400

O

- one-hot encoding, 57
- online resources, 402
- operating environment, 1
- ordinal value, 56
- outer product, 5
- outliers, 80
- overfitting, 94, 194, 207

P

- p-value, 8
- parent distribution, 7, 61, 246
- partial derivative, 196
- Pasteur, Louis, 166
- Pillow, 3, 49
 - convert, 50
 - fromarray, 50
 - open, 50
 - save, 50
 - show, 50
 - documentation, 50
- Plato, 7, 61
- pooling, 292

- positive predictive value, 252
- power spectrum, 383
- PPV, 252
- precision, 252
- precision-recall curve, 269
- preprocessing, 83
- primary visual cortex, 290
- principal component analysis (PCA),
 - 95, 153, 157
 - MNIST, 157
- prior class probability, 62, 246
- probability, 6
 - Bayes' theorem, 113
 - distribution, 7
 - Gaussian distribution, 8
 - joint probability, 115
 - likelihood, 114
 - normal distribution, 7, 8, 116
 - notation
 - uniform distribution, 8
 - parent distribution, 7, 61
 - posterior probability, 114
 - prior class probability, 62
 - prior probability, 114
 - uniform distribution, 7
- probability distribution, 7
- probability map, 329
- pseudorandom sequence, 139
- Python
 - blocks, 12
 - bottom-tested loop, 21
 - break, 22
 - continue, 22, 23
 - control structures, 19
 - for, 20
 - if, 12, 19
 - try, 23
 - while, 21
 - with, 23
 - data structures, 13
 - debugging, 24
 - dictionaries, 13, 18
 - documentation, 27
 - dynamic typing, 14
 - editor conventions, 13
 - enumerate, 20
 - essence, 11
 - except, 23
 - exceptions, 23
 - exiting, 12
 - False, 16
 - floating-point numbers, 13
 - for loops, 20
 - function definition, 24
 - defaults, 25
 - if statement, 19
 - indentation, 12
 - interactive mode, 12
 - librosa, 370
 - list comprehensions, 21
 - list operations, 15
 - lists, 13, 15
 - copying, 16
 - lists in memory, 30
 - module, 25, 26
 - copy, 17
 - deepcopy, 17
 - time, 26, 32
 - namespaces, 26
 - None, 16
 - pass, 13
 - pickle, 46, 181, 308
 - primitive data types, 15
 - range, 20
 - statements, 12
 - strings, 13, 14
 - quotes, 15
 - top-tested loop, 21
 - True, 16
 - try, 23
 - tuple, 17
 - variables, 13, 14
 - camel-case, 14
 - naming, 14
 - while loop, 21, 22
 - white space, 12
 - with statement, 23
- Python Crash Course* (Matthes), 11
- PyTorch, 2

R

- random forest, 122, 163
- Random Numbers and Computers, 48
- recall, 251
- receiver operating characteristics, 259
- receptive field, 291
- rectified linear unit (relu), 172
- recurrent neural networks, 402

- recursion, 119
- Reddit Machine Learning, 402
- regularization, 94, 186, 207, 210
 - dropout, 212
 - L2, 210, 234
 - weight decay, 211, 234
- reinforcement learning, 400
 - AlphaGo, 401
 - AlphaGo Zero, 401
 - Atari 2600, 401
- RMSprop, 307, 320
- RNN, 402
- ROC, 259
 - elements of, 262, 263

S

- sample, 4
- scalar, 5
- scikit-learn, 2
 - DecisionTreeClassifier, 130
 - documentation, 2
 - GaussianNB, 130
 - KNeighborsClassifier, 130
 - make_classification, 71
 - MLPClassifier, 170, 183, 218
 - MultinomialNB, 130
 - NearestCentroid, 130
 - RandomForestClassifier, 130
 - SVC, 130
- SciPy, 370
 - wavfile, 370
- semantic segmentation, 400
- sensitivity, 251
- SGD, 190, 220, 320, 338
- sigmoid, 172
- softmax, 175
- specificity, 251
- spectrogram, 384
- standard deviation, 6
- standard error, 6
- standardizing, 66, 67, 135
- statistical test, 8
- statistically significant, 8
- statistics, 6
 - bootstrap sample, 122
 - descriptive, 6, 7
 - dot product, 45
 - 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, 80
- standard deviation, 6, 33, 66, 116
- standard error, 6, 33, 77
- statistically independent, 114
- statistically significant, 8
- t-test, 8
- variance, 6
- stochastic gradient descent, 190, 220, 307, 320
- supervised learning, 59, 400
- Support Vector Machine (SVM), 164
 - kernel, 126
 - margin, 123, 125
 - one-versus-one, 155
 - one-versus-rest, 155
 - optimizing, 126
 - support vector, 125

T

- t-SNE, 300
- t-test, 8
- tangent line, 188
- tensor, 6
- TensorFlow, 2
 - documentation, 2
- test data, 69
- TNR, 251
- toolkit
 - Caffe, 2
 - Caffe2, 2
 - HDF5, 3
 - installation, 3
 - Keras, 2
 - MXnet, 2
 - NumPy, 2
 - Pillow, 3
 - PyTorch, 2
 - scikit-learn, 2
 - TensorFlow, 2
 - versions, 3
- TPR, 251
- training data, 69

transcendental functions, 171
transfer learning, 352
 embedding, 353
trigonometric functions, 171
true negative rate, 251
true positive rate, 251

U

U-Net, 333
Ubuntu, 1
UCI Machine Learning Repository, 84
uniform distribution, 7
unsupervised learning, 400

V

validation data, 69
variance, 6
vector, 4, 31
 column vector, 4
 dot product, 5
 inner product, 5
 multiplication, 5
 outer product, 5
 row vector, 4

W

weight decay, 211
weight initialization, 186, 205, 237
 Glorot, 206
 He, 206
 Xavier, 206
weight update, 189, 190
weighted accuracy, 273
weights, 170
Windows, 3

Y

YOLO, 333
Youden's J statistic, 255
YouTube, 403

Z

zero-padding, 280