

Index

A

ALSO, 240
ALSO-E, 240
Alternative clustering, 2
Autoencoder ensembles, 14, 134
Average k -nearest neighbor detector, 214
Average of maximum (AOM) combination, 144

B

Bagging, 107
Ball-and-marble data set, 265
Ball-and-speck data set, 116, 265
Base model choice, 7
Benchmarking dilemmas, 267
Bias (Intuitive explanation), 36
Bias reduction, 163
Bias (theoretical definition), 41
Boosting, 166
Bragging, 137
BSS, 24
Bucket of models, 27

C

Calibrated bagging, 19
CARE, 24
Categorization of ensemble algorithms, 6
Computational variance, 79

D

Damped averaging, 22
Data-centered ensembles, 14
Data-centric variance reduction, 86
DBSS, 24

Deterministic pruning, 171
Distance distribution-based outliers, 227

E

Ensembles, 9
ERC-forest, 2, 101

F

Feature bagging, 94
Feature engineering, 181
Filtered variable probability sampling (FVPS), 24, 174
Fisher's score, 116, 268
Fixed bias sampling, 172

G

GASP, 243
Geometric subsampling (GS), 134
Geometric subsampling with rotated bagging (GR), 136
Gravity-defying behavior, 115

H

Harmonic k -nearest neighbor detector, 216
HiCS, 19, 97
High-dimensional ensembles, 5
High-dimensional outlier detection, 94
HS-trees, 106

I

iForest, 102, 221
Independent ensemble pseudocode, 11

Independent ensembles, 9, 11
 Instability in variance reduction, 86
 Intrusion bootstrap, 17
 Isolation forests, 102, 221

K

Kernel density methods, 224
 Kernel fisher discriminant, 237
 Kernel Mahalanobis method, 231
 Kernel whitening, 231

L

Limitations of variance reduction, 84
 Local outlier factor (LOF), 217
 LODA, 24, 100

M

Mahalanobis distance, 227
 Mahalanobis method, 226
 Many could be better than all, 168
 Maximum of average (MOA) combination, 145
 Model-centered ensembles, 12
 Model-centric variance reduction, 91
 Model pruning, 168
 Model selection, 4
 Multiple-Proclus, 14, 19, 101, 220
 Multiview clustering, 2

N

Netflix prize, 3
 No-Free-Lunch theorem, 83
 Normalization of outlier scores, 8
 Nyström approximation to kernel Mahalanobis, 236

O

One-class SVM, 237
 OutRank, 19, 100, 220
 OUTRES, 19, 24, 97
 Overfitting, 39

P

Precision, 191
 PROCLUS, 14, 101, 220

R

RandNet, 24
 Random forests, 25
 Randomized feature weighting (RFW), 136
 Recall, 191
 Receiver operating characteristic (ROC), 191
 Relevance weighted ensemble, 181
 Representational bias, 80
 Representational variance, 79
 Rotated bagging (RB), 19, 99
 RS-Hash, 19, 24, 102, 220
 RS-stream, 102, 221

S

Sapling forests, 106
 SCiforest, 106
 SELECT, 19, 175
 Sequential ensemble pseudocode, 9
 Sequential ensembles, 8
 Stacking, 2, 26, 182
 Statistical variance, 79
 Subbagging, 107
 Subsampling, 19, 107
 Subspace histograms, 220
 Subspace outlier detection, 178
 Support-vector data description (SVDD), 238

T

Thresh combination, 144
 Training data pruning, 167, 171
 TRINITY, 247

V

Variable bagging (VB), 126
 Variable bias sampling, 174
 Variable subsampling (VS), 119
 Variable subsampling with rotated bagging (VR), 124
 Variance (Intuitive explanation), 36
 Variance reduction, 75
 Variance (theoretical definition), 41

W

WAG, 130
 Wagging, 130