

# Index

- $\chi^2$  distance, 144
- active contours, *see* edge detection
- active learning, 70, 197
- adaptive mesh refinement, 70, 136
- agglomerative algorithms, *see* clustering
- ambiguity in problem definition, 55
- angles, *see* features, problem-specific
- angular radial transform, *see* features, shape
- anomaly detection, *see* pattern recognition
- aperture problem, 132, 156
- association rules, 201–202
- axis-parallel trees, 186
  
- background subtraction, *see* moving object detection
- backpropagation algorithm, 192
- backward elimination, *see* wrappers, feature selection
- bagging, 194
- bilateral filter, *see* filters, image processing
- bipartite graph matching, 203
- blessings of dimensionality, 175
- blind deconvolution, *see* image contrast enhancement
- blind source separation, 166
- block-matching techniques, *see* moving object detection
- boosting, 195
- box plots, *see* visualization
  
- Canny edge detector, *see* edge detection
- CART-LC method, 187
- chord distribution, *see* features, shape
- city-block distance, *see* distance metrics
- class separability filter, *see* filters, feature selection
  
- classification, 184–199
  - accuracy, 196
  - cross validation, 196
  - decision trees, 186
  - ensembles, 194
  - evaluation data, 196
  - in region detection, 127
  - $k$ -nearest neighbor, 185
  - model interpretability, 184
  - naïve Bayes, 185
  - neural networks, 189
  - SVMs (support vector machines), 193
  - test data, 184
  - training data, 184
- closing, *see* mathematical morphology
- clustering, 178–184
  - agglomerative algorithms, 180
  - choice of cluster centers, 183
  - cluster validity, 183
  - clustering criterion, 179
  - clustering tendency, 183
  - complete linkage, 180
  - divisive algorithms, 180
  - graph algorithms, 181
  - in region detection, 127
  - $k$ -means, 179
  - $k$ -medioids, 180
  - number of clusters, 183
  - similarity metric, 179
  - single linkage, 180
- code emulators, *see* code surrogates
- code surrogates, 29, 199
- coherent structures, 25, 28
- competitive learning algorithms, 168
- complete linkage, *see* clustering
- compressed sensing, 167

- computer simulations, 25
  - clustering, 181
  - mesh data, 43, 136
- concept drift, 199
- concept indexing, *see* dimension reduction
- confidence of a rule, 202
- conjugate gradient, *see* mathematical optimization
- connected component labeling, 136
- contour shape descriptor, *see* features, shape
- contrast stretching, *see* image contrast enhancement
- corner density, *see* features, problem-specific
- corner detector
  - Harris, 129
  - Moravec, 129
  - USAN (Univalue Segment Assimilating Nucleus), 129
- correlation filter, *see* filters, feature selection
- cosine distance, *see* distance metrics
- cross validation, *see* classification
- curse of dimensionality, 62, 81, 91, 162
  
- data fusion, 79–92
  - catastrophic fusion, 81
  - decision level, 82, 91–92
  - feature level, 82, 90–91
  - JDL (Joint Directors of Laboratories) model, 81
  - sensor level, 81, 83–89
- data mining, *see* scientific data mining
- deblurring, *see* image contrast enhancement
- decision trees, *see* classification
- Dempster–Shafer theory, 92
- descriptors, *see* features
- design of experiments, 29, 70
- diffusion equation, 102
- dilation, *see* mathematical morphology
- dimension
  - definition, 161
- dimension reduction
  - concept indexing, 173
  - domain-specific methods, 172–174
  - feature selection methods, 168–172
  - feature transformation methods, 164–168
  - incremental methods, 173
- dimension reduction, feature selection
  - filters, 169
  - regression, 172
  - wrappers, 171
- dimension reduction, feature transforms
  - EOFs (empirical orthogonal functions), 164
  - FastMap, 167
  - Hotelling transform, 164
  - independent component analysis, 166
  - Karhunen-Loève transform, 164
  - kernel PCA (principal component analysis), 166
  - latent semantic indexing, 164
  - locally linear embedding, 166
  - multidimensional scaling, 167
  - POD (proper orthogonal decomposition), 164
  - PCA (principal component analysis), 164
  - principal curves, 166
  - projection pursuit, 166
  - random projections, 166
  - self-organizing maps, 168
  - SVD (singular value decomposition), 164
- distance metrics, 205
  - city-block distance, 205
  - cosine distance, 205
  - Euclidean distance, 205
  - Minkowski distance, 205
  - Tanimoto distance, 205
- divisive algorithms, *see* clustering
- domain decomposition, 181
  
- earth-mover’s distance, 144
- edge density, *see* features, problem-specific
- edge detection, 114–123
  - active contours, 117
  - Canny edge detector, 117
  - Marr–Hildreth technique, 115
  - Prewitt operator, 114

- Roberts' operator, 114
- salient edges, 138
- Sobel operator, 114
- USAN (Univalve Segment Assimilating Nucleus) approach, 123
- edge relaxation, 138
- Eikonal equation, 121
- ensemble learning, 91
- ensembles, *see* classification
- EOFs (Empirical Orthogonal Functions), *see* dimension reduction, feature transforms
- erosion, *see* mathematical morphology
- Euclidean distance, *see* distance metrics
- evaluation data, *see* classification
- evolutionary algorithms, *see* mathematical optimization
- exhaustive search, *see* wrappers, feature selection
- expectation maximization, 180
  
- FastMap, *see* dimension reduction, feature transforms
- feature selection methods, *see* dimension reduction
- features
  - definition, 51, 141
  - postprocessing, 157–159
  - problem-specific, 153–157
  - requirements, 142
  - shape, 146–149
  - simple, 144–146
  - texture, 149–153
- features, postprocessing
  - derived features, 159
  - missing features, 158
  - normalization, 158
  - outlier identification, 158
  - units, 159
- features, problem-specific, 153–157
  - angles, 153
  - edge-corner density, 155
  - graph-based, 154
  - SIFT (Scale-Invariant Feature Transform) features, 155
  - velocity determination, 156
- features, shape, 146–149
  - ART (Angular Radial Transform), 147
  - chord distribution, 147
  - contour shape descriptor, 148
  - Fourier descriptors, 146
  - shape distribution, 147
  - shapelets, 148
  - skeleton, 147
- features, simple, 144–146
  - geometric moments, 145
  - histograms, 144
  - location of centroid, 145
  - scalar descriptors, 145
  - simple statistics, 144
- features, texture, 149–153
  - Gabor filter, 152
  - grey-level cooccurrence matrix, 150
  - power spectrum, 151
  - wavelets, 152
- filters, feature selection
  - class separability filter, 169
  - correlation filter, 170
  - PCA (principal component analysis) filter, 170
  - Relief, 170, 172
  - stump filter, 170
- filters, image processing
  - bilateral filter, 98
  - discrete convolution, 95
  - Gaussian filter, 96
  - gradient inverse weighted filter, 98
  - homomorphic filtering, 106
  - low-pass filter, 96
  - mean filter, 95
  - median filter, 99
  - MMSE (minimum mean-squared error) filter, 96
  - Prewitt operator, 114
  - Roberts' operator, 114
  - sharpening, 108
  - Sobel operator, 114
  - USAN (Univalve Segment Assimilating Nucleus) filter, 98
  - Weiner filter, 109
- forward selection, *see* wrappers, feature selection

- Fourier descriptors, *see* features, shape
- frame differencing, 132
- Gabor filter, *see* features, texture
- Gaussian filter, *see* filters, image processing
- Gaussian mixture models, 133
- Gini split criterion, 188
- GIS (Graphic Information Systems), 17
- gradient descent, *see* mathematical optimization
- gradient inverse weighted filter, *see* filters, image processing
- granulometry, 110
- graph Laplacian, 182
- graph partitioning  
in clustering, 182  
in segmentation, 127
- graph-based features, *see* features, problem-specific
- grey-level cooccurrence matrix, *see* features, texture
- Hilditch's thinning algorithm, 147
- histogram equalization, *see* image contrast enhancement
- histograms, 144  
comparison of, 144, 169
- hit-or-miss transform, *see* mathematical morphology
- homomorphic filtering, 106
- Hotelling Transform, *see* dimension reduction, feature transforms
- Hu moments, 7, 145
- hypergraph, 182
- hysteresis thresholding, 117
- image contrast enhancement, 107–109  
blind deconvolution, 109  
contrast stretching, 107  
deblurring, 109  
histogram equalization, 108  
multiscale contrast enhancement, 108  
sharpening, 108
- image denoising, 95–107  
filter-based approaches, 95–99  
morphological methods, 110–111  
multiplicative noise, 105–106  
PDE (partial differential equation)–based approaches, 102–105  
problem-specific approaches, 106–107  
wavelet-based approaches, 99–102
- image noise  
Gaussian, 95  
multiplicative, 95, 105  
salt-and-pepper, 95
- image processing software  
CVIPtools, 222  
Fv, 215  
GIMP, 214, 222  
Image, 222  
ImageMagick, 214, 222  
OpenCV, 222
- image registration, 82, 83, 85–89, 128
- image segmentation  
cleanup, 138  
domain-specific techniques, 135–136  
edge-based techniques, 114–123  
moving object detection, 132–135  
object identification, 136–138  
object representation, 138  
region-based techniques, 123–128  
salient regions, 128–132
- incremental algorithms, 199
- incremental learning, 91
- independent component analysis, *see* dimension reduction, feature transforms
- indexing, 163, 174, 185
- information gain ratio split criterion, 188
- information gain split criterion, 188
- information retrieval, 27, 34, 90, 163
- information visualization, 38, 163
- Isomap, 167
- itemset, 202
- Johnson–Lindenstrauss lemma, 166
- k*-means, *see* clustering
- k*-medioids, *see* clustering
- k*-nearest neighbor, *see* classification; *see* regression
- Kalman filter, 133, 204

- Karhunen–Loève transform, *see* dimension reduction, feature transforms
- KDD, *see* knowledge discovery in databases
- kernel PCA (principal component analysis), *see* dimension reduction, feature transforms
- knowledge discovery in databases, 2
- Kullback–Leibler divergence, 144, 169
  
- Latent Semantic Indexing, *see* dimension reduction, feature transforms
- lazy learning, 185
- Legendre moments, 146
- levels sets, 119
- Levenberg–Marquardt method, *see* mathematical optimization
- linear discriminant function, 190
- linear regression, *see* regression
- locally linear embedding, *see* dimension reduction, feature transforms
- locally weighted regression, *see* regression
  
- Marr–Hildreth technique, *see* edge detection
- MARS (multivariate adaptive regression splines), 200
- matched filter method, 135
- mathematical morphology, 110, 128, 138, 147
  - closing, 110
  - dilation, 110
  - erosion, 110
  - hit-or-miss transform, 110
  - opening, 110
- mathematical optimization
  - conjugate gradient, 206
  - evolutionary algorithms, 88, 206
  - gradient descent, 88, 206
  - Levenberg–Marquardt method, 88, 206
  - simulated annealing, 88, 206
- max minority split criterion, 189
- mean filter, *see* filters, image processing
- medial axis transform, 147
- median absolute deviation, 98, 102
- median filter, *see* filters, image processing
- metamodels, *see* code surrogates
- MMSE (minimum mean-squared error) filter, *see* filters, image processing
- Minkowski distance, *see* distance metrics
- misclassification cost, 197
- missing features, *see* features, postprocessing
- model trees, *see* regression
- moments, *see* features, simple
- Morse–Smale complex, 128
- moving object detection, 132
  - background subtraction, 132
  - block-matching techniques, 133
- multidimensional scaling, *see* dimension reduction, feature transforms
- multiresolution
  - contrast enhancement, 108
  - Gaussian pyramid, 72, 96, 134
  - Laplacian pyramid, 73, 90
  - wavelets, 73
  
- naïve Bayes, *see* classification
- neural networks, *see* classification
- nonmaximal suppression, 117
- normalization, *see* features, postprocessing
  
- oblique trees, 187
- opening, *see* mathematical morphology
- optical flow, 89, 156
- outlier detection, *see* pattern recognition
  
- parallel plots, *see* visualization
- particle filter, 204
- pasting, 195
- pattern recognition
  - anomaly detection, 204
  - association rules, 201–202
  - classification, 184–199
  - clustering, 178–184
  - distance metrics, 205
  - mathematical optimization, 206
  - outlier detection, 204
  - regression, 199–201
  - tracking, 202–204

- pattern recognition software
  - R, 222
  - Weka, 222
- PCA (principal component analysis), *see*
  - dimension reduction, feature transforms
- PCA filter, *see* filters, feature selection
- PDE (partial differential equation)–based
  - techniques
  - denoising, 102
  - edge detection, 119
- Pearson’s correlation coefficient, 170
- perceptron, 190
- pixel connectivity
  - 4-connectivity, 136
  - 6-connectivity, 136
  - 8-connectivity, 136
  - 26-connectivity, 136
- POD (Proper Orthogonal Decomposition),
  - see* dimension reduction, feature transforms
- power spectrum, *see* features, texture
- Prewitt operator, *see* edge detection
- principal curves, *see* dimension reduction, feature transforms
- projection pursuit, *see* dimension reduction, feature transforms
- projection pursuit regression, 193
- projects
  - ADaM (Algorithm Development and Mining System), 224
  - Diamond Eye, 9, 223
  - FOCAS (Faint Object Classification and Analysis), 6
  - JARTool (JPL Adaptive Recognition Tool), 9
  - Quakefinder, 14
  - Sapphire, 9, 224
  - SKICAT (Sky Imaging Cataloging and Analysis Tool), 8
- pruning, 189
- random forests, 195
- random projections, *see* dimension reduction, feature transforms
- region detection, 123–132
  - classification, 127
  - clustering, 127
  - region merging, 124
  - region splitting, 124
  - region splitting and merging, 125
  - thresholding, 123
  - watershed algorithm, 128
- region merging, *see* region detection
- region splitting, *see* region detection
- region splitting and merging, *see* region detection
- registration, *see* image registration
- regression, 199–201
  - ensembles, 201
  - k*-nearest neighbor, 200
  - linear regression, 200
  - locally weighted regression, 201
  - MARS (multivariate adaptive regression splines), 200
  - model trees, 201
  - neural networks, 201
  - regression trees, 201
  - SVMs (support vector machines), 201
- regression trees, *see* regression
- Relief, *see* filters, feature selection
- reordering schemes, 182
- response surface modeling, 29
- Retinex method, 106
- Roberts’ operator, *see* edge detection
- salient regions
  - corners, 129
  - edges, *see* edge detection
  - scale saliency, 129
  - Scale-Invariant Feature Transform, 131
- sampling
  - adaptive, 70
  - iterative, 69
  - random, 68
  - stratified, 69
  - systematic, 68
- scale saliency, *see* salient regions
- scale-space theory, 73, 96
- scatter plots, *see* visualization

- scientific data, 4
- scientific data mining
  - analysis characteristics, 53–55
  - applications, 6–39
  - data characteristics, 46–53
  - data types, 41–46
  - definition, 1, 65
  - end-to-end process, 57–66
  - projects, *see* projects
- segmentation, *see* image segmentation
- self-organizing maps, *see* dimension reduction, feature transforms
- semisupervised learning, 197
- shape distribution, *see* features, shape
- shapelets, *see* features, shape
- sharpening, *see* image contrast enhancement
- SIFT (scale-invariant transform theory), *see* salient regions
- SIFT features, *see* features, problem-specific
- sigmoidal unit, 192
- simulated annealing, *see* mathematical optimization
- single linkage, *see* clustering
- SVD (Singular Value Decomposition), *see* dimension reduction, feature transforms
- skeleton, *see* features, shape
- sketch, 167
- snakes, 118
- Sobel operator, *see* edge detection
- software
  - ARPACK, 166, 222
  - CLUTO, 184
  - CVIPtools, 222
  - GIMP, 222
  - GMV, 222
  - Gnuplot, 222
  - hMETIS, 182
  - Image, 222
  - ImageMagick, 222
  - LAPACK, 166, 222
  - METIS, 182
  - OpenCV, 222
  - OpenGL, 222
  - R, 222
  - SCALAPACK, 166, 222
  - SVDPACK, 166, 222
  - Tcl/Tk, 222
  - VisIt, 222
  - Weka, 222
- spectral graph theory, 182
- split criterion, 187
- streaming data, 167
- structural saliency, 138
- stump filter, *see* filters, feature selection
- sum minority split criterion, 189
- support of a rule, 202
- surrogate models, *see* code surrogates
- SVMs (support vector machines), *see* classification
- Tanimoto distance, *see* distance metrics
- tensor voting, 138
- test data, *see* classification
- thresholding, *see* region detection
- tracking, 84, 202–204
  - multiple hypothesis algorithm, 84
  - probabilistic data association, 84
- training data, *see* classification
- twoing rule split criterion, 188
- unbalanced training set, 198
- uncertainty quantification, 54
- USAN (Univalve Segment Assimilating Nucleus) approach, *see* edge detection
- USAN filter, *see* filters, image processing
- vector quantization, 180
- velocity determination, *see* features, problem-specific
- virtual observatories, 12
- visualization
  - box plots, 210
  - image data, 214
  - mesh data, 214
  - parallel plots, 212
  - scatter plots, 211
- visualization software
  - GGobi, 213
  - GMV, 214, 222
  - Gnuplot, 210, 222
  - OpenGL, 222
  - RoSuDa, 213

- Tcl/Tk, 222
- VisIt, 214, 222
- XmdvTool, 213
- Voronoi diagrams, 180, 185
- watershed algorithm, *see* region detection
- wavelets
  - data fusion, 90
  - decimated transform, 76, 100
  - denoising, 99
  - registration, 88
  - texture features, 152
  - undecimated transform, 76, 100
- Weiner filter, 109, 133
- wrappers, feature selection
  - backward elimination, 171
  - exhaustive search, 171
  - forward selection, 171
- Zernike moments, 146