
Index

- $\sqrt{\text{Gini}}$, *see* impurity measure, $\sqrt{\text{Gini}}$
- 0–1 loss, **62**
- 0-norm, **234**
- 1-norm, **234**, **239**
- 2-norm, **234**, **239**

- abstraction, **306**
- accuracy, **19**, **54**, **57**
 - as a weighted average
 - Example 2.1, **56**
 - expected
 - Example 12.3, **347**
 - Example 12.1, **346**
 - macro-averaged, **60**
 - ranking, *see* ranking accuracy
- active learning, **122**, **361**
- adjacent violators, **77**
- affine transformation, **195**
- agglomerative merging, **312**
- AggloMerge(S, f, Q)
 - Algorithm 10.2, **312**
- aggregation
 - in structured features, **306**
- Aleph, *see* ILP systems
- analysis of variance, **355**
- anti-unification, **123**

- Apriori, *see* association rule algorithms
- AQ, *see* rule learning systems
- arithmetic mean
 - minimises squared Euclidean distance, **291**
 - Theorem 8.1, **238**
- association rule, **15**, **184**
- association rule algorithms
 - Apriori, **193**
 - Warmr, **193**
- association rule discovery
 - Example 3.12, **102**
- AssociationRules(D, f_0, c_0)
 - Algorithm 6.7, **185**
- at least as general as, **105**
- attribute, *see* feature
- AUC, **67**
 - multi-class
 - Example 3.5, **88**
- average recall, *see* recall, average, **101**

- backtracking search, **133**
- bag of words, **41**
- bagging, **156**, **331–333**
- Bagging(D, T, \mathcal{A})
 - Algorithm 11.1, **332**

- basic linear classifier, **21**, 207–228, **238**,
 241, 242, 269, 317, 321, 332, 334–
 336, 339, 364
 - Bayes-optimality, 271
 - kernel trick, 43
- Bayes' rule, **27**
- Bayes-optimal, **29**, **30**, **265**
 - basic linear classifier, 271
- beam search, **170**
- Bernoulli distribution, 148, **274**
 - multivariate, **273**
- Bernoulli trial, 148, **273**
- Bernoulli, Jacob, 45, 274
- BestSplit-Class(D, F)
 - Algorithm 5.2, 137
- bias, **94**
- bias–variance dilemma, **93**, 338
- big data, 362
- bigram, 322
- bin, **309**
- binomial distribution, **274**
- bit vector, **273**
- Bonferroni–Dunn test, **357**
- boosting, 63, 334–338
 - weight updates
 - Example 11.1, 334
- Boosting(D, T, \mathcal{A})
 - Algorithm 11.3, 335
- bootstrap sample, **331**
- breadth-first search, 183
- Brier score, **74**

- C4.5, *see* tree learning systems
- calibrating classifier scores, 223, 316
- calibration, 220
 - isotonic, **78**, 223, 286, **318**
 - logistic, 286, **316**
 - loss, **76**
 - map, **78**
- CART, *see* tree learning systems

- Cartesian product, **51**
- categorical distribution, **274**
- CD, *see* critical difference
- central limit theorem, **220**, 350
- central moment, **303**
- centre around zero, 24, **198**, **200**, 203,
 324
- centre of mass, 24
- centroid, **97**, **238**
- characteristic function, **51**
- Chebyshev distance, **234**
- Chebyshev's inequality, **301**
- Chervonenkis, Alexey, 125
- chicken-and-egg problem, 287
- cityblock distance, *see* Manhattan dis-
 tance
- class
 - imbalance
 - Example 2.4, 67
 - label, **52**
 - ratio, 57, 58, 61, 62, 67, 70, 71, 142,
 143, 147, 221
- class probability estimation, 360
- class probability estimator, **72**
 - squared error
 - Example 2.6, 74
 - tree, 141
- classification
 - binary, **52**
 - multi-class, **14**, 82
- classifier, **52**
- clause, **105**
- cloud computing, 362
- clustering, **14**
 - agglomerative, **255**, 310
 - descriptive, **18**
 - evaluation
 - Example 3.10, 99
 - predictive, **18**

- representation
 - Example 3.9, 98
 - stationary point, **249**
 - Example 8.5, 249
- clustering tree, 253
 - using a dissimilarity matrix
 - Example 5.5, 153
 - using Euclidean distance
 - Example 5.6, 155
- CN2, *see* rule learning systems
- CNF, *see* conjunctive normal form
- comparable, **51**
- complement, 178
- component, **266**
- computational learning theory, 124
- concavity, **76**
- concept, 157
 - closed, **117**, 183
 - complete, **113**
 - conjunctive, 106
 - Example 4.1, 106
 - consistent, **113**
- concept learning, **52**, **104**
 - negative examples
 - Example 4.2, 110
- conditional independence, 281
- conditional likelihood, **283**
- conditional random field, 296
- confidence, 57, **184**
- confidence interval, **351**
 - Example 12.5, 352
- confusion matrix, **53**
- conjugate prior, **265**
- conjunction \wedge , **34**, **105**
- conjunctive normal form, **105**, 119
- conjunctively separable, **115**
- constructive induction, **131**
- contingency table, **53**
- continuous feature, *see* feature, quanti-
 - tative
- convex, **213**, **241**
 - hull, **78**
 - lower, **309**
 - loss function, **63**
 - ROC curve, **76**, 138
 - set, **113**, **183**
- correlation, 151
- correlation coefficient, **45**, **267**, 323
- cosine similarity, **259**
- cost ratio, **71**, 142
- count vector, **275**
- counter-example, **120**
- covariance, **45**, 198
- covariance matrix, **200**, 202, 204, 237,
 - 267, 269, 270
- coverage counts, **86**
 - as score
 - Example 3.4, 87
- coverage curve, **65**
- coverage plot, **58**
- covering algorithm, **163**
 - weighted, *see* weighted covering
- covers, **105**, **182**
- critical difference, **356**
- critical value, **354**
- cross-validation, 19, **349**
 - Example 12.4, 350
 - internal, **358**
 - stratified, **350**
- curse of dimensionality, **243**
- d-prime, **316**
- data mining, 182, **362**
- data set characteristics, 340
- data streams, **361**
- De Morgan laws, **105**, 131
- decile, **301**
- decision boundary, 4, 14
- decision list, **34**, 192

- decision rule, **26**
- decision stump, **339**
- decision threshold
 - tuning
 - Example 2.5, 71
- decision tree, **33**, 53, 101, 314
- decoding, **83**
 - loss-based, **85**
- deduction, **20**
- deep learning, **361**
- default rule, **35**, **161**
- degree of freedom
 - in t -distribution, **353**
 - in contingency table, **58**
- dendrogram, **254**
 - Definition 8.4, 254
- descriptive clustering, **96**, 245
- descriptive model, **17**
- dimensionality reduction, 326
- Dirichlet prior, **75**
- discretisation, 155
 - agglomerative, **310**
 - agglomerative merging
 - Example 10.7, 313
 - bottom-up, **310**
 - divisive, **310**
 - equal-frequency, **309**
 - equal-width, **310**
 - recursive partitioning
 - Example 10.6, 311
 - top-down, **310**
- disjunction \vee , **34**, **105**
- disjunctive normal form, **105**
- dissimilarity, 96, **152**
 - cluster, **152**
 - split, **152**
- distance, **23**
 - elliptical
 - Example 8.1, 237
 - Euclidean, **23**, 305
 - Manhattan, **25**
- distance metric, **235**, 305
 - Definition 8.2, 236
- distance weighting, **244**
- divide-and-conquer, 35, **133**, 138, 161
- DKM, **156**
- DNF, *see* disjunctive normal form
- dominate, **59**
- DualPerceptron(D)
 - Algorithm 7.2, 209
- Eddington, Arthur, 343
- edit distance, **235**
- eigendecomposition, **325**
- Einstein, Albert, 30, 343
- EM, *see* Expectation-Maximisation
- empirical probability, **75**, 133, 135, 138
- entropy, 159, **294**
 - as an impurity measure, *see* impurity measure, entropy
- equivalence class, **51**
- equivalence oracle, **120**
- equivalence relation, **51**
- error rate, **54**, 57
- error-correcting output codes, 102
- estimate, **45**
- Euclidean distance, **234**
- European Conference on Machine Learning, 2
- European Conference on Principles and Practice of Knowledge Discovery in Databases, 2
- evaluation measures, 344
 - for classifiers, 57
- example, **50**
- exceptional model mining, 103
- excess kurtosis, *see* kurtosis
- exemplar, **25**, **97**, 132, **238**
- Expectation-Maximisation, 97, **289**, 322

- expected value, **45**, 267
- experiment, **343**
- experimental objective, **344**
- explanation, **35**
- explanatory variable, *see* feature
- exponential loss, **63**, 337
- extension, **105**

- F-measure, **99**, **300**
 - insensitivity to true negatives, 346
- false alarm rate, **55**, 57
- false negative, **55**
- false negative rate, **55**, 57
- false positive, **55**
- false positive rate, **55**, 57
- feature, **13**, **50**, 262
 - binarisation, **307**
 - Boolean, **304**
 - calibration, **314**
 - categorical, 155, **304**
 - construction, **41**, 50
 - decorrelation, 202, 237, 270, 271
 - discretisation, **42**, **309**
 - discretisation, supervised, **310**
 - discretisation, unsupervised, **309**
 - domain, **39**, **50**
 - list, **33**
 - normalisation, 202, 237, 270, 271, **314**
 - ordinal, 233, **304**
 - quantitative, **304**
 - space, **225**
 - structured, **306**
 - Example 10.4, 306
 - thresholding, **308**
 - thresholding, supervised, **309**
 - thresholding, unsupervised, **308**
 - transformation, **307**
 - two uses of, 41
 - Example 1.8, 41
 - unordering, **307**
- feature calibration, 277
 - categorical
 - Example 10.8, 315
 - isotonic
 - Example 10.11, 321
 - Example 10.10, 320
 - logistic
 - Example 10.9, 318
- feature selection, 243
 - backward elimination, **324**
 - filter, **323**
 - forward selection, **324**
 - Relief, **323**
 - wrapper, **324**
- feature tree, **32**, **132**, 155
 - Definition 5.1, 132
 - complete, **33**
 - growing
 - Example 5.2, 139
 - labelling
 - Example 1.5, 33
- first-order logic, **122**
- FOIL, *see* ILP systems
- forecasting theory, **74**
- frequency, *see* support
- FrequentItems(D, f_0)
 - Algorithm 6.6, 184
- Friedman test, **355**
 - Example 12.8, 356
- function estimator, **91**
- functor, **189**

- Gauss, Carl Friedrich, 102, 196
- Gaussian distribution, **266**
- Gaussian kernel, **227**
 - bandwidth, **227**
- Gaussian mixture model, 266, **289**
 - bivariate
 - Example 9.3, 269

- relation to K -means, 292
 - univariate
 - Example 9.2, 269
- general, **46**
- generalised linear model, 296
- generality ordering, **105**
- generative model, **29**
- geometric median, **238**
- Gini coefficient, **134**
- Gini index, 159
 - as an impurity measure, *see* impurity measure, Gini index
- Gini, Corrado, 134
- glb, *see* greatest lower bound
- Golem, *see* ILP systems
- Gosset, William Sealy, 353
- gradient, **213**, 238
- grading model, 52, 92
- Gram matrix, **210**, 214, 325
- greatest lower bound, **108**
- greedy algorithm, **133**
- grouping model, 52, 92
- GrowTree(D, F), 152
 - Algorithm 5.1, 132
- Guinness, 353

- HAC(D, L)
 - Algorithm 8.4, 255
- Hamming distance, **84**, **235**, 305
- harmonic mean, 99
- Hernández-Orallo, José, xvi
- hidden variable, **16**, **288**
- hierarchical agglomerative clustering, 314
- hinge loss, **63**, **217**
- histogram, **302**
 - Example 10.2, 303
- homogeneous coordinates, 4, **24**, 195, 201
- Horn clause, **105**, 119, **189**
- Horn theory, 119
 - learning
 - Example 4.5, 122
- Horn(Mb, Eq)
 - Algorithm 4.5, 120
- Horn, Alfred, 105
- Hume, David, 20
- hyperplane, **21**
- hypothesis space, **106**, 186

- ID3, *see* tree learning systems
- ILP, *see* inductive logic programming
- ILP systems
 - Aleph, 193
 - FOIL, 193
 - Golem, 193
 - Progol, 35, 193
- implication \rightarrow , **105**
- impurity
 - Example 5.1, 136
 - relative, **145**
- impurity measure, 158
 - $\sqrt{\text{Gini}}$, **134**, 145, 147, 156, 337
 - entropy, **134**, 135, 136, 144, 147, 294
 - Gini index, **134**, 135, 136, 144, 147
 - as variance, 148
 - minority class, **134**, 135, 159
- imputation, **322**
- incomparable, **51**
- independent variable, *see* feature
- indicator function, **54**
- induction, **20**
 - problem of, xvii, **20**
- inductive bias, **131**
- inductive logic programming, **189**, 307
- information content, **293**
 - Example 9.7, 293
- information gain, **136**, 310, 323
- information retrieval, 99, 300, 326, 346
- input space, **225**
- instance, **49**

- labelled, **50**
- instance space, **21**, 39, 40, **49**
 - segment, **32**, 51, 104, **132**
- intercept, **195**, 198
- internal disjunction, **110**, 162
 - Example 4.3, 112
- interquartile range, **301**, 314
- isometric
 - $\sqrt{\text{Gini}}$, 145
 - accuracy, 61, 69, 77, 116
 - average recall, **60**, 72
 - entropy, 145
 - Gini index, 145
 - impurity, 159
 - precision, 167
 - precision (Laplace-corrected), 170
 - splitting criteria, 145
- item set, **182**
 - closed, **183**
 - frequent, **182**
- Jaccard coefficient, **15**
- jackknife, 349
- K -means, 25, 96, 97, **247**, 259, 289, 310
 - problem, **246**, 247
 - relation to Gaussian mixture model, 292
- K -medoids, **250**, 310
- k -nearest neighbour, **243**
- Karush–Kuhn–Tucker conditions, **213**
- kernel, **43**, 323
 - quadratic
 - Example 7.8, 225
- kernel perceptron, **226**
- kernel trick, **43**
 - Example 1.9, 44
- Kernel-KMeans(D, K)
 - Algorithm 8.5, 259
- KernelPerceptron(D, κ)
 - Algorithm 7.4, 226
- Kinect motion sensing device, 129, 155
- KKT, *see* Karush–Kuhn–Tucker conditions
- KMeans(D, K)
 - Algorithm 8.1, 248
- KMedoids(D, K, Dis)
 - Algorithm 8.2, 250
- kurtosis, **303**
- L_0 norm, *see* 0-norm
- label space, **50**, 360
- Lagrange multiplier, **213**
- landmarking, 342
- Langley, Pat, 359
- Laplace correction, **75**, 138, 141, 147, 170, 265, 274, 279, 286
- lasso, **205**
- latent semantic indexing, **327**
- latent variable, *see* hidden variable
- lattice, **108**, 182, 186
- law of large numbers, **45**
- learnability, **124**
- learning from entailment, **128**
- learning from interpretations, **128**
- learning model, **124**
- learning rate, **207**
- LearnRule(D), 169, 190
 - Algorithm 6.2, 164
- LearnRuleForClass(D, C_i)
 - Algorithm 6.4, 171
- LearnRuleList(D), 169, 190
 - Algorithm 6.1, 163
- LearnRuleSet(D)
 - Algorithm 6.3, 171
- least general generalisation, **108**, 112, 115–117, 131
- least upper bound, **108**
- least-squares classifier, **206**, 210
 - univariate
 - Example 7.4, 206

- least-squares method, **196**
 - ordinary, **199**
 - total, **199**, 273
- least-squares solution to a linear regression problem, 272
- leave-one-out, **349**
- Lebowski, Jeffrey, 16
- level-wise search, 183
- Levenshtein distance, **235**
- LGG, *see* least general generalisation
- LGG-Conj(x, y)
 - Algorithm 4.2, 110
- LGG-Conj-ID(x, y)
 - Algorithm 4.3, 112
- LGG-Set(D)
 - Algorithm 4.1, 108
- lift, **186**
- likelihood function, **27**, 305
- likelihood ratio, **28**
- linear
 - approximation, **195**
 - combination, **195**
 - function, **195**
 - model, **194**
 - transformation, **195**
- linear classifier, 5, 21, 38, 40, 43, 81, 82, 207–223, 263, 282, 314, 333–340
 - Example 1, 3
 - coverage curve, 67
 - general form, 207, 364
 - geometric interpretation, 220
 - logistic calibration
 - Example 7.7, 223
 - margin, 22
 - VC-dimension, 126
- linear discriminants, 21
- linear regression, 92, 151
 - bivariate
 - Example 7.3, 203
 - univariate
 - Example 7.1, 198
- linear, piecewise, **195**
- linearly separable, 207
- linkage function, **254**
 - Definition 8.5, 254
 - Example 8.7, 256
 - monotonicity, **257**
- literal, **105**
- Lloyd's algorithm, 247
- local variables, **189**, **306**
- log-likelihood, 271
- log-linear models, **223**
- log-odds space, 277, **317**
- logistic function, **221**
- logistic regression, 223, **282**
 - univariate
 - Example 9.6, 285
- loss function, **62**, 93
- loss-based decoding
 - Example 3.3, 86
- L_p norm, *see* p -norm
- LSA, *see* latent semantic indexing
- lub, *see* least upper bound
- m -estimate, **75**, 141, 147, **279**
- Mach, Ernst, 30
- machine intelligence, 361
- machine learning
 - definition of, 3
 - univariate, **52**
- Mahalanobis distance, **237**, 271
- majority class, **33**, 35, **53**, **56**
- Manhattan distance, **234**
- manifold, **243**
- MAP, *see* maximum a posteriori
- margin
 - of a classifier, 62
 - of a decision boundary, **211**
 - of a linear classifier, **22**

- of an example, **62**, 211, 339
- margin error, **216**
- marginal, **54**, **186**
- marginal likelihood, **29**
 - Example 1.4, 30
- market basket analysis, 101
- matrix
 - diagonal, 202
 - inverse, 267
 - rank, **326**
- matrix completion, **327**
- matrix decomposition, 17, 97, 324–327
 - Boolean, 327
 - non-negative, 328
 - with constraints, 326
- maximum a posteriori, **28**, 263
- maximum likelihood, **28**
- maximum-likelihood estimation, **271**, 287
 - in linear regression, 199
- maximum-margin classifier
 - Example 7.5, 216
 - soft margin
 - Example 7.6, 219
- mean, 267, **299**
 - arithmetic, **300**
 - geometric, **300**
 - harmonic, **300**
- mean squared error, **74**
- median, 267, **299**
- medoid, 153, **238**
- membership oracle, **120**
- meta-model, **339**
- MGConsistent(C, N)
 - Algorithm 4.4, 116
- midrange point, **301**
- minimum description length
 - Definition 9.1, 294
- Minkowski distance, **234**
 - Definition 8.1, 234
- minority class
 - as an impurity measure, *see* impurity measure, minority class
- missing values, 322
 - Example 1.2, 27
- mixture model, **266**
- ML, *see* maximum likelihood
- MLM data set
 - Example 1.7, 40
 - clustering
 - Example 8.4, 249
 - hierarchical clustering
 - Example 8.6, 254
- mode, 267, **299**
- model, **13**, **50**
 - declarative, **35**
 - geometric, **21**
 - grading, **36**
 - grouping, **36**
 - logical, 32
 - parametric, **195**
 - probabilistic, 25
 - univariate, **40**
- model ensemble, **330**
- model selection, **265**
- model tree, **151**
- monotonic, **182**, **305**
- more general than, **105**
- MSE, *see* mean squared error
- multi-class classifier
 - performance
 - Example 3.1, 82
- multi-class probabilities
 - Example 3.7, 91
- multi-class scores
 - from decision tree, 86
 - from naive Bayes, 86
 - reweighting
 - Example 3.6, 90

- multi-label classification, **361**
- multi-task learning, **361**
- multinomial distribution, **274**
- multivariate linear regression, **277**
- multivariate naive Bayes
 - decomposition into univariate models, **32**
- multivariate normal distribution, **289**
- multivariate regression
 - decomposition into univariate regression, **203, 364**
- n*-gram, **322**
- naive Bayes, **30, 33, 203, 278, 315, 318, 322**
 - assumption, **275**
 - categorical features, **305**
 - diagonal covariance matrix, **281**
 - factorisation, **277, 281**
 - ignores feature interaction, **44**
 - linear in log-odds space, **277**
 - multi-class scores, **86**
 - prediction
 - Example 9.4, **276**
 - recalibrated decision threshold, **277, 365**
 - Scottish classifier, **32, 281**
 - skewed probabilities, **277**
 - training
 - Example 9.5, **280**
 - variations, **280**
- nearest-neighbour classifier, **23, 242**
- nearest-neighbour retrieval, **243**
- negation \neg , **105**
- negative recall, **57**
- neighbour, **238**
- Nemenyi test, **356**
- neural network, **207**
- Newton, Isaac, **30**
- no free lunch theorem, **20, 340**
- noise, **50**
 - instance, **50**
 - label, **50**
- nominal feature, *see* feature, categorical
- normal distribution, **220, 266, 305**
 - multivariate, **267**
 - multivariate standard, **267**
 - standard, **267, 270**
- normal vector, **195**
- normalisation, **198**
 - row, **90**
- null hypothesis, **352**
- objective function, **63, 213**
- Occam's razor, **30**
- one-versus-one, **83**
- one-versus-rest, **83**
- online learning, **361**
- operating condition, **72**
- operating context, **345**
- optimisation
 - constrained, **212, 213**
 - dual, **213, 214**
 - multi-criterion, **59**
 - primal, **213**
 - quadratic, **212, 213**
- Opus, *see* rule learning systems
- ordinal feature, *see* feature, ordinal
- ordinal, **299**
- outlier, **198, 238, 302**
 - Example 7.2, **199**
- output code, **83**
- output space, **50, 360**
- overfitting, **19, 33, 50, 91, 93, 97, 131, 151, 196, 210, 285, 323**
 - Example 2, **6**
- p*-norm, **234**
- p*-value, **352**

Index

393

- PAC, *see* probably approximately correct
- paired t -test, **353**
 Example 12.6, 353
- PAM, *see* partitioning around medoids
- PAM(D, K, Dis)
 Algorithm 8.3, 251
- Pareto front, **59**
- partial order, **51**
- partition, **51**
- partition matrix, **97**
- partitioning around medoids, **250, 310**
- PCA, *see* principal component analysis
- Pearson, Karl, 299
- percentile, **301**
 Example 10.1, 302
- percentile plot, **301**
- perceptron, **207, 210**
 online, **208**
- Perceptron(D, η)
 Algorithm 7.1, 208
- PerceptronRegression(D, T)
 Algorithm 7.3, 211
- piecewise linear, *see* linear, piecewise
- population mean, **45**
- post-hoc test, **356**
- post-processing, **186**
- posterior odds, **28**
 Example 1.3, 29
- posterior probability, **26, 262**
- powerset, **51**
- Príncipe, 343
- precision, **57, 99, 167, 186, 300**
 Laplace-corrected, 170
- predicates, *see* first-order logic
- predicted positive rate, **347**
- predictive clustering, **96, 245, 289**
- predictive model, **17**
- predictor variable, *see* feature
- preference learning, **361**
- principal component analysis, 24, 243,
 324
- prior odds, **28**
- prior probability, **27**
- probabilistic model
 discriminative, **262**
 generative, **262**
- probability distribution
 cumulative, **302**
 right-skewed, **303**
- probability estimation tree, 73–76, 141,
 147, 262, 263, 265
- probability smoothing, **75**
- probability space, **317**
- probably approximately correct, **124, 331**
- Progol, *see* ILP systems
- projection, 219
- Prolog, *see* query languages
- propositional logic, **122**
- propositionalisation, **307**
- PruneTree(T, D)
 Algorithm 5.3, 144
- pruning, **33, 142**
- pruning set, **143**
- pseudo-counts, **75, 172, 274, 279**
- pseudo-metric, **236**
- pure, **133**
- purity, 35, 158
- quantile, **301**
- quartile, **301**
- query, **306**
- query languages
 Prolog, 189–191, 193, 306
 SQL, 306
- Rand index, **99**
- random forest, 129, **333, 339**
- random variable, **45**
- RandomForest(D, T, d)

- Algorithm 11.2, 333
- range, **301**
- ranking, **64**
 - Example 2.2, 64
 - accuracy
 - Example 2.3, 65
- ranking accuracy, **64**
- ranking error, **64**
- ranking error rate, **64**
- recalibrated likelihood decision rule, **277**
- recall, 57, **58**, 99, 300
 - average, **60**, **178**
- receiver operating characteristic, **60**
- RecPart(S, f, Q)
 - Algorithm 10.1, 311
- recursive partitioning, **310**
- reduced-error pruning, **143**, 147, 151
 - incremental, 192
- refinement, **69**
- refinement loss, **76**
- regression, **14**, 64
 - Example 3.8, 92
 - isotonic, 78
 - multivariate, **202**
 - univariate, **196**
- regression coefficient, **198**
- regression tree
 - Example 5.4, 151
- regressor, **91**
- regularisation, **204**, 217, 294
- reinforcement learning, **360**
- reject, **83**
- relation, **51**
 - antisymmetric, **51**
 - equivalence, *see* equivalence relation
 - reflexive, **51**
 - symmetric, **51**
 - total, **51**
 - transitive, **51**
- residual, **93**, **196**
- ridge regression, **205**
- Ripper, *see* rule learning systems
- ROC curve, **67**
- ROC heaven, **69**, 145
- ROC plot, **60**
- Rocchio classifier, **241**
- rotation, **24**
- rule, **105**
 - body, **157**
 - head, **157**
 - incomplete, **35**
 - inconsistent, **35**
 - overlap
 - Example 1.6, 35
- rule learning systems
 - AQ, 192
 - CN2, 192, 357
 - Opus, 192
 - Ripper, 192, 341
 - Slipper, 341
 - Tertius, 193
- rule list, 139, **157**
 - Example 6.1, 159
 - as ranker
 - Example 6.2, 165
- rule set, **157**
 - Example 6.3, 167
 - as ranker
 - Example 6.4, 173
- rule tree, **175**
 - Example 6.5, 175
- sample complexity, **124**
- sample covariance, **45**
- sample mean, **45**
- sample variance, **45**
- scale, **299**
 - reciprocal, **300**

- scaling, **24**
 - uniform, **24**
- scaling matrix, **201**
- scatter, **97, 246**
 - Definition 8.3, **246**
 - reduction by partitioning
 - Example 8.3, **247**
 - within-cluster, **97**
- scatter matrix, **200, 245, 325**
 - between-cluster, **246**
 - within-cluster, **246**
- scoring classifier, **61**
- Scottish classifier, *see* naive Bayes
- SE, *see* squared error
- search heuristic, **63**
- seed example, **170**
- segment, **36**
- semi-supervised learning, **18**
- sensitivity, **55, 57**
- separability
 - conjunctive
 - Example 4.4, **116**
- separate-and-conquer, **35, 161, 163**
- sequence prediction, **361**
- sequential minimal optimisation, **229**
- set, **51**
 - cardinality, **51**
 - complement, **51**
 - difference, **51**
 - disjoint, **51**
 - intersection, **51**
 - subset, **51**
 - union, **51**
- Shannon, Claude, **294**
- shatter, **126**
- shattering a set of instances
 - Example 4.7, **126**
- shrinkage, **204**
- sigmoid, **221**
- signal detection theory, **60, 316**
- significance test, **352**
- silhouette, **252**
- similarity, **72**
 - Example 1.1, **15**
- singular value decomposition, **324**
- skewness, **303**
 - Example 10.3, **304**
- slack variable, **216, 294**
- Slipper, *see* rule learning systems
- slope, **195**
- soft margin, **217**
- SpamAssassin, **1–14, 61, 72**
- sparse data, **22**
- sparsity, **205**
- specific, **46**
- specificity, **55, 57**
- split, **132**
 - binary, **40**
- splitting criterion
 - cost-sensitivity
 - Example 5.3, **144**
- SQL, *see* query languages
- squared error, **73**
- squared Euclidean distance, **238**
- stacking, **339**
- standard deviation, **301**
- statistic
 - of central tendency, **299**
 - of dispersion, **299**
 - shape, **299, 303**
- stop word, **280**
- stopping criterion, **163, 310**
- structured output prediction, **361**
- sub-additivity, **236**
- subgroup, **100, 178**
 - evaluation
 - Example 6.6, **180**
 - extension, **100**

- subgroup discovery, 17
 - Example 3.11, 100
- subspace sampling, 333
- supervised learning, 14, 17, 46
- support, 182
- support vector, 211
- support vector machine, 22, 63, 211, 305
 - complexity parameter, 217
- SVD, *see* singular value decomposition
- SVM, *see* support vector machine

- t*-distribution, 353
- target variable, 25, 91, 262
- task, 13
- terms, *see* first-order logic
- Tertius, *see* rule learning systems
- test set, 19, 50
- Texas Instruments TI-58 programmable
 - calculator, 228
- text classification, 7, 11, 22
- thresholding
 - Example 10.5, 309
- total order, 51
- training set, 14, 50
- transaction, 182
- transfer learning, 361
- translation, 24
- tree learning systems
 - C4.5, 156
 - CART, 156
 - ID3, 155
- triangle inequality, 236
- trigram, 322
- true negative, 55
- true negative rate, 55, 57
- true positive, 55
- true positive rate, 55, 57, 99
- turning rankers into classifiers, 278

- underfitting, 196

- unification, 123
 - Example 4.6, 123
- unigram, 322
- universe of discourse, 51, 105
- unstable, 204
- unsupervised learning, 14, 17, 47

- Vapnik, Vladimir, 125
- variance, 24, 45, 94, 149, 198, 200, 301, 303
 - Gini index as, 148
- VC-dimension, 125
 - linear classifier, 126
- version space, 113
 - Definition 4.1, 113
- Viagra, 7–44
- vocabulary, 7
- Voronoi diagram, 98
- Voronoi tessellation, 241
- voting
 - one-versus-one
 - Example 3.2, 85

- Warmr, *see* association rule algorithms
- weak learnability, 331
- weighted covering, 181, 338
 - Example 6.7, 181
- weighted relative accuracy, 179
- WeightedCovering(*D*)
 - Algorithm 6.5, 182
- Wilcoxon's signed-rank test, 354
 - Example 12.7, 355
- Wilcoxon-Mann-Whitney statistic, 80

- χ^2 statistic, 101, 312, 323
- Xbox game console, 129

- z*-score, 267, 314, 316