

Index

- α -helix, 26
 - biochemical structure, 28
- β -strand, 26
- χ^2 -distribution, 9
- 4-optim, *see* algorithms, 4-optim
- 5-optim, *see* algorithms, 5-optim
- AA, *see* amino acids
- active site, 26
- additive tree, *see* phylogenetic tree construction
- adjusted currency, 16
- algorithms
 - 4-optim, 173
 - 5-optim, 174
 - analytic, 116
 - Brent minimization, 202
 - CHE procedures, 148, 188, 190
 - discrete steepest descent (SD), **53**
 - improved version, 54
 - early abort, **52**
 - evolutionary (EVA), 185
 - genetic, 185
 - crossover event, 188
 - mutation and recombination rates, 190
 - rotation, 188
 - swapping, 188
 - greedy algorithms, 50
 - NNSearch, 76
 - PiecewiseConstMax, **118**, 119
 - random directions, 209
 - refining a linear classification, 63
 - UPGMA, 166
 - WPGMA, 168
- amino acids, **25**
- approximation
 - best, 3
 - continuous, 50
 - direction of least confidence, 12
 - linear, **18**
 - piecewise linear, 16, **23**
 - total error, **34**
- ask price (stock market), 97
- balanced tree, 72
- barrier, 59
- BB, *see* best basis
- beacon, 1
 - DME, distance measuring equipment, 1
 - VOR, very high frequency omnirange, 1
- best basis, 49, 108
- bid price (stock market), 97
- binary search tree, 70
- book of shares (stock market), 97
- Brent
 - exponential search, 202
 - method, **198**
 - minimization, 202
- bucketing, 79
- building NN trees, **74**
- car replacement, *see* replacement, of cars
- Cauchy–Schwartz inequality, 76
- charge, 62
- CHE, *see* algorithms, CHE procedures
 - optimization strategy, 148
 - procedures, 188, 190
 - sensitivity analysis, 191
- chi-square distribution, 9
- classification, 63
- close price (stock market), 97
- cluster analysis, 80
- clustering, 80
 - bad dimension, **80**
 - good dimension, 80
 - large gap versus median, 81
 - unbalancing the tree, 81
- complexity of the solution space, 55

- computer replacement, *see* replacement of
 - computers
- concanavalin A, 26
- confidence analysis, 8, 124
 - center of gravity, 124
 - direction of least confidence, 12
 - Monte Carlo estimation, 125
 - smoothness of the optimal area, 125
 - transfer of error, 125
 - volume of the optimal solution, 125
 - with random columns, 58
- confidence interval, 56
- constructor–heuristic–evaluator, *see* algorithms,
 - CHE procedures
- cost
 - of operation, 16
 - of repairs function, 16
- covariance matrix, 72, 192
 - not singular, 79
- criteria for discarding singular values, 40, **43**, 45
- cytochrome c_2 precursor, 27
- data points, 67
 - closest, 67
- decision problem, 32
- deletion, 153, **161**
- diagonal matrix, 39
- direction of most confidence, 12
- discrete steepest descent (SD), *see* algorithms,
 - discrete steepest descent
- distance based methods, *see* phylogenetic tree
 - construction, distance based methods
- distance function, 67
- DME, *see* beacon, distance measuring equipment
- dynamic programming, 127
 - backtracking, 137
 - examples, 129
 - optimal binary search trees, 129
 - optimal order of matrix multiplications, 132
 - string matching, 135
 - scheme, 128
- early abort (EA), *see* algorithms, early abort
- eigenvalue, 39
 - decomposition, 39
 - eigenvector of smallest eigenvalue, 12
- eigenvector/eigenvalue decomposition, 12, 39
- ellipse, 12
- embarrassingly parallel, 190
- error
 - norm of error, 4
 - normalized error vector, 5
 - total, 3
- error analysis, 8
- euclidean norm, 35
- EVA, *see* algorithms, evolutionary
- EVD, *see* eigenvalue decomposition
- evolution, 140
 - algorithms, *see* algorithms, evolutionary
- exponential decay, 15
- exponential weighted average, 78
- false negative and false positive, **60**, 62
- Fisher’s linear discriminant, 62
- function minimization, **196**
 - methods (overview), 197
 - multivariate methods
 - analytical method, 205
 - Brent in a given direction, 208
 - comparison of methods, 220
 - Newton’s method, 212
 - random directions, 208
 - spectral method, 214
 - steepest descent, 207
 - univariate methods
 - analytical method, 197
 - Brent’s algorithm, 198
 - Brent’s initial exponential search, 202
 - combined approach, 205
 - golden section search, 198
 - parabolic interpolation, 204
- GA, *see* algorithms, genetic
- genetic algorithms, *see* algorithms, genetic
- greedy algorithm, *see* algorithms, greedy
- Hessian matrix, 11
- Hessian normal form, 71
- hydrophobicity, 61
- hyperplane, 71, **72**
 - bad choice, 73
 - good choice, 73
- imbalance of information, 35
- indel, 161
- inflation, 16
- informative character, *see* phylogenetic tree
 - construction, character based methods

insertion, 153, 161	symmetric, 39
interpolation phase, 77	transformation, 36
 	measuring errors, 4
<i>k</i> -dimensional tree or <i>k</i> -d-tree, 70	minimizing a function subject to constraints, 6
 	modelling, <i>see</i> mathematical modelling
learning methods, 66	molecular weight, 61
least squares, 25	Moore's law, 22
non-linear, 61	multidimensional maximization methods
on exponentials, 20	adaptive algorithms, 123
problem, 4, 5	confidence analysis, 124
singular value decomposition,	grid search, 123
33	random directions, 120
with best basis, 49	shrinking hypercube, 122
leghemoglobin, 26	spiral search, 121
lexicographical order, 34	
limit order (stock market), 98	near singular problems, 41
linear approximation, 18	nearest neighbor tree, <i>see</i> NN, trees
linear programming, 82, 93–94,	nearest neighbors, 66
110	negative edge length, 172
comparison with least squares,	neighbor, 51
92	NMR, <i>see</i> nuclear magnetic resonance
prediction, 92	NN, 71
slack variables, 91	problem, 69
linguistics, 141	trees, 68
local optimum, 54	building of, 74
LP, <i>see</i> linear programming	searching in, 75
LS, <i>see</i> least squares	NNSearch, 76
 	node, 71
mapping, 67	of a NN tree, 72
market order (stock market), 98	non-linear least squares problem, 5
Markovian model, 153	solution, 6
mathematical modelling, 30	normal distribution, 4, 192
computer program as model, 31	normalization, 67
independent sets of data, 31	normalized error vector, 5
linear programming, 31	normalizing dimensions, 79
neural network optimization, 31	NP-complete, 49, 62, 91
prediction accuracy, 30	nuclear magnetic resonance (NMR), 27
test data, 30	
use of test data, 31	OBST, <i>see</i> optimal binary search trees
training data, 30, 67	Occam's razor, 176
training the model, 30	one-dimensional structure, 25
validation, 30	open price (stock market), 97
matrix	optimal binary search trees, 129
covariance, 72	optimal transaction sequence (OTS), <i>see</i> stock
diagonal, 36, 39	market prediction, optimal transaction
Hessian, 11	sequence
notation, 34, 62	optimum, local, 54
orthonormal, 36, 39	order (stock market), 98
positive definite, 12	order price (stock market), 97
singular, 79	orthonormal matrix, 39

- OTS, *see* stock market prediction, optimal transaction sequence
- outlier, 63
- PAM
 - 1-PAM mutation matrix, 153
 - 250-PAM mutation matrix, 155
 - distance, 152, **153**, 170, 172
 - mutation matrix, 154
- parabolic interpolation, 204
- parallelism, 190
- path, 53
- path signature, 119
- peptide bond, 30
- perfect phylogeny, 181
- phylogenetic tree, 139
- phylogenetic tree construction, 139, 142
 - character based methods, 144, 175–181
 - general parsimony, 177
 - informative character, 175
 - perfect phylogeny, 181
 - strict character compatibility, 181
 - CHE methods, 148
 - classification types, 143
 - creating an initial topology, 166
 - Dayhoff matrices, 158
 - definition of PAM distance, 156
 - deletion, 161
 - determining the root with an outlier, 147
 - distance based methods, 144, 149–175
 - measures of distance, 149
 - PAM distance, 152
 - percentage of change, 151
 - synteny, 152
 - examples
 - string matching, 143
 - improving trees
 - 4-optim, 173
 - 5-optim, 174
 - swapping heuristics, 173
 - indel, 161
 - insertion, 161
 - model of evolution, 153
 - number of trees, 146
 - parsimony, 176
 - probabilistic methods, 144
 - rooted trees, 147
 - sequence alignment, 160
 - scoring insertions/deletions, 161
 - tree building, 145
 - tree fitting by least squares, 171
 - triosephosphate isomerase, 144
 - unrooted tree, 146
 - UPGMA algorithm, 166
 - WPGMA algorithm, 168
- piecewise linear approximation, *see* approximation, piecewise linear
- piecewise linear functions, 23
- positive definite matrix, 12
- prediction, 67
 - of secondary structures, 25
- primary structure, 25
- principal value, 5
- problem
 - least squares, 4
 - near singular, 41
 - non-linear least squares, 5
 - overdetermined, 3, 34
 - singular, 40
- protein, **25**
 - biochemistry, **25**
- quad trees, 70
- quadratic form, 11
- random columns, 58
- random directions (RD), 79, 208
- random starting point, 54
- randomness, 190
- range searching, 69
- RD, *see* random directions
- refining
 - a linear classification, 63
 - the predictor function, 59
- relation between SVD and EVD, **40**
- replacement
 - of cars, 15
 - of computers, 22
 - paradox, 14
- residual error, 45, 51
- residue
 - insignificant contribution, 41
 - minimal size, 36
- ribbon, 27

searching <i>k</i> nearest neighbors, 76 - NN trees, 75 - NNSearch, 76	fractal nature, 104
secondary structure, 26	interactive exercise, 102
secondary structure prediction, 25 - as modelling problem, 32 - using linear programming, 89	least squares, 110
sensitivity analysis for parallel CHE, 191	modelling, 100
sequence alignment, 160 - cost-free end alignment, 164 - distance and variance matrices, 165 - estimating distances using maximum likelihood, 162 - global alignment, 164 - local alignment, 164 - scoring insertions/deletions, 161	numerical problem, 104, 107
sequential search, 69	optimal transaction sequence (OTS), 101
shrinking hypercube, <i>see</i> multidimensional maximization methods	period data, 104, 107
simulation, 111 - optimization of simulation parameters, 114 - simulating decision processes, 115	simulation, 111
singular matrix, 79	stock market terms, 97
singular problem, 40	stop order (stock market), 98
singular value decomposition, 25, 33 , 36 , 43 - actual calculation, 36 - online calculation, 37	strict character compatibility, 181
singular values, 36 criteria for discarding, 40, 43, 45	structure - concanavaline A, 26 - cytochrome <i>c</i>₂ precursor, 27 - leghemoglobin, 26 - primary, 25 - secondary, 26 - tertiary, 26
slack variables, 86	subtree, 76
sliding window, 33	SVD, <i>see</i> singular value decomposition
spiral search, <i>see</i> multidimensional maximization methods	swapping heuristics, <i>see</i> phylogenetic tree construction, improving trees
spread (stock market), 97	symmetric matrix, 39
SSP, <i>see</i> secondary structure prediction	system of equations, 4, 34
standard deviation, 4	
steepest descent, 207 improved algorithm, 54	tangent function, 5 - inverting the function, 5 - principal value, 5
stepwise regression, 49	taxonomy, 140
stock market prediction, 96 - approximating the OTS, 103 - computing optimal parameters, 109 - concepts, 97 - decision problem, 104 - dimensionality of the decision function, 108	Taylor series, 10
	tertiary structure, 26
	test data, <i>see</i> mathematical modelling, test data
	threshold value, 71
	ticker symbol (stock market), 98
	total error of approximation, 34
	training data, <i>see</i> mathematical modelling, training data
	training the model, <i>see</i> mathematical modelling, training the model
	transaction (stock market), 97
	transformation matrix, 36
	tree - additive tree, 185 - balanced, 72 - binary search tree, 70 - Indo-European languages, 142 <i>k</i>-d tree, 70 - NN tree, 72 - building of, 74 - searching in, 75

tree (<i>cont.</i>)	VOR, <i>see</i> beacon, very high frequency
node, 71	omnirange
quad tree, 70	
searching <i>k</i> nearest neighbors, 76	weighted average, 78
subtree, 76	weighted exponential, 78
tree of life, 139	weighted function, 61
ultrametric tree, 184	weighted number of errors, 60
	weighting, 35
UPGMA, <i>see</i> algorithms, UPGMA	WPGMA, <i>see</i> algorithms, WPGMA
validation, 60, 67	X-ray crystallography, 27
variance, 107, 168, <i>169</i> , 192	
largest, 72	Z-score, <i>see</i> Z-transform, 79
volume (stock market), 97	Z-transform, 79