
Index

- active set methods, 136
- adaboost, 77
- Adatron, 25, 134
- affine function, 81
- ANOVA kernels, 50

- Bayes formula, 75
- bias, 10, 130
- bioinformatics, 157
- box constraint algorithm, 108

- capacity, 56
 - control, 56, 93, 112
- chunking, 136
- classification, 2, 9, 93
- compression, 69, 100, 112
- computational learning theory, 77
- computer vision, 152
- consistent hypothesis, 3, 54
- constraint
 - active and inactive, 80
 - box, 107
 - equality and inequality, 80
- convex
 - combination, 166
 - function, 80
 - hull, 89
 - programming, 81
 - set, 80
- convex optimisation problem, 81
- covariance matrix, 48
- cover, 59
- covering number, 59
- curse of dimensionality, 28, 63

- data dependent analysis, 58, 70, 78

- decision function, 2
- decomposition, 136
- dimension
 - fat-shattering, *see* fat-shattering dimension
 - VC, *see* VC dimension
- dimensionality reduction, 28
- distribution free, 54
- dot product, 168
- dual
 - optimisation problem, 85
 - problem, 85
 - representation, 18, 24, 30
 - variables, 18
- duality, 87
 - gap, 86

- effective VC dimension, 62
- eigenvalues, eigenvectors, 171
- empirical risk minimisation, 58
- error of a hypothesis, 53

- fat-shattering dimension, 61
- feasibility gap, 99, 109, 127
- feasible region, 80, 126, 138
- feature space, 27, 46
- features, 27, 36, 44
- flexibility, 56

- Gaussian
 - kernel, 44
 - prior, 48
 - process, 48, 75, 120, 144
- generalisation, 4, 52
- gradient ascent, 129
- Gram matrix, 19, 24, 41, 169

- growth function, 55
- hand-written digit recognition, 156
- hard margin, 94
- Hilbert space, 36, 38, 169
- hinge loss, 77, 135
- hyperplane, 10
 - canonical, 94
- hypothesis space, 2, 54
- image classification, 152
- implicit mapping, 30
- inner product, 168
- input distribution, 53
- insensitive loss, 112
- insensitive loss functions, 114
- Karush–Kuhn–Tucker (KKT) conditions, 87, 126
- kernel, 30
 - Adatron, 134
 - making, 32
 - matrix, 30
 - Mercer, 36
 - reproducing, 39
- Krieger, 119
- Kuhn–Tucker theorem, 87
- Lagrange
 - multipliers, 83
 - theorem, 83
 - theory, 81
- Lagrangian, 83
 - dual problem, 85
 - generalised, 85
- learning
 - algorithm, 2
 - bias, 6, 93, 112
 - machines, 8
 - methodology, 1
 - probably approximately correct, *see* pac learning
 - rate, 129
 - theory, 52
- least squares, 21
- leave-one-out, 101, 123
- linear learning machines, 9
- Linear Programme, 80
- linear programming SVMs, 112
- linearly separable, 11
- Lovelace, Lady, 8
- luckiness, 69
- margin
 - bounds on generalisation, 59
 - distribution, 12, 59
 - error, 64, 107
 - functional, 12, 59
 - geometric, 12
 - slack variable, 15, 65
 - soft, 65
 - soft and hard, 69
- matrix
 - positive definite, 171
 - positive semi-definite, 171
- maximal margin
 - bounds, 59
 - classifier, 94
- maximum a posteriori, 75
- Mercer's
 - conditions, 34
 - theorem, 33
- model selection, 53, 101, 120
- multi-class classification, 123
- neural networks, 10, 26
- noisy data, 65
- nonseparable, 11
- Novikoff's theorem, 12
- objective
 - dual, 85
 - function, 80, 125
- on-line learning, 3, 77
- operator
 - linear, 171
 - positive, 35
- optimisation problem
 - convex, *see* convex optimisation problem
- optimisation theory, 79
- pac learning, 54

- perceptron, 10, 135
- primal
 - form, 11
 - problem, 80
- quadratic optimisation, 96, 98, 106, 108, 117, 118
- quadratic programming, 80, 88, 125, 135
- recursive kernels, 44, 50
- regression, 20, 112
- reproducing
 - kernel Hilbert space, 38
 - property, 39
- ridge regression, 22, 118
- risk functional, 53
- saddle point, 86
- sample complexity, 53
- sample compression scheme, 69, 112
- scalar product, 168
- sequential minimal optimization (SMO), 137
- shattering, 56, 61
- Skilling, 144
- slack
 - variables, 65, 71, 80, 104
 - vector, 65, 72, 104
- soft margin
 - algorithm, 1-norm, 107
 - algorithm, 2-norm, 105
 - bounds, 65
 - optimisation, 102
- sparseness, 100, 114
- statistical learning theory, 77
- stopping criteria, 127
- structural risk minimisation, 58
 - data dependent, 62, 78
- supervised learning, 1
- Support Vector Machines, 7, 93
- support vectors, 97
- target function, 2
- text categorisation, 150
- Turing, 8
- uniform convergence, 55
- unsupervised learning, 3
- value of an optimisation problem, 80, 85
- VC dimension, 56
- VC theorem, 56
- Widrow–Hoff algorithm, 22, 145
- working set methods, 136