
Index

- Active learning, 304
- AI; *see* Artificial intelligence
- AI problem set, 325
- Approximation
 - ϵ -exact, 177
 - low rank, 47, 232
 - multipole, 47, 175
 - sparse, 47, 206, 231
- Artificial intelligence, 323

- Basis selection, 257
- Bayesian committee machine, 213
- BCM; *see* Bayesian committee machine
- BLAS, 113
- Blind source separation, 225

- CCCP; *see* Concave-convex procedure
- Chunking, 13
- COCO; *see* Constrained covariance
- Communication overhead, 122
- Comparing classifiers, 145
- Concave-convex procedure, 276, 279
- Conjugate
 - directions, 15, 45
 - gradient, 45, 159
- Constrained covariance, 229
- Convex optimization, 6, 275
- Convolutional networks, 319, 348
- Core vector machine, 142
- Covariance function, 204
- Covariance operator, 228
- Curse of dual variables, 277
- CVM; *see* Core vector machine

- DC; *see* Difference of convex

- Decomposition method, 16, 83, 107, 141
- Deformations, 311, 350
- Deterministic annealing, 164
- Difference of convex, 276
- Direction search, 13, 305
- Dual optimization, 1, 105, 139
- Dual-tree methods, 180
- Duality, 6
 - strong, 8
 - weak, 8

- Evidence, 219

- Farthest point clustering, 186
- Fast Gauss transform, 181, 189
- Fast matrix-vector multiplication,
 - see* Matrix-vector multiplication
- Fast multipole methods, 47, 175, 181
- Feasible direction, 13, 305
- Feature selection, 67
- FGT; *see* Fast Gauss transform

- Gauss transform, 179
- Gaussian process, 176, 204
- Genomic sequence analysis, 73
- GP; *see* Gaussian process
- Gradient
 - conjugate, 45, 159
 - descent
 - on orthogonal group, 236
 - stochastic, 349
 - projection, 14

- Hash table, 79
- Hessian, 14, 256

- Hilbert-Schmidt independence
 - criterion, 230
- Hinge loss, 37, 278
- HSIC; *see* Hilbert-Schmidt
 - independence criterion
- Huber loss, 37
- ICA; *see* Independent component analysis
- IFGT; *see* Improved fast Gauss transform
- Improved fast Gauss transform, 181, 189
- Incomplete Cholesky decomposition, 26, 47, 232
- Independence measure, 228
- Independent component analysis, 225, 227
- Inducing
 - inputs, 207
 - variables, 207, 218
- Interior point methods, 141
- Invariance, 301, 308, 348
- Inverted index, 51
- Iterative chunking, 13
- k -mers, 75, 98
- Karush Kuhn Tucker, 6, 280
- KDE; *see* Kernel density estimation
- Kernel, 4
 - cache, 12, 25, 66, 116, 259
 - computation, 11
 - independence measure, 228
 - inverted index, 51
 - linear, 157
 - local, 337
 - string kernel; *see* String kernel
- Kernel density estimation, 176, 195
- Kernel independent component analysis, 225
- Kernel matching pursuit, 254, 264
- KKT; *see* Karush Kuhn Tucker
- KMP; *see* Kernel matching pursuit
- Lagrangian duality, 6
- LAPACK, 113
- Large-scale learning, 2
- LASVM, 306
 - example selection, 307
 - finishing, 315
 - process, 307
 - reprocess, 307
- Learning
 - active, 304
 - large-scale, 2
 - online, 303
 - semisupervised, 161, 285
 - small-scale, 2
- Learning architectures
 - deep, 332, 345
 - shallow
 - fixed preprocessing, 330
 - template matching, 330
 - trainable features, 331
- LIBSVM, 20, 27, 43, 69, 118
- Line search, 160, 237
- List merging, 58
- LLAMA, 69
- Local learning, 337
- LOQO, 27
- Loss
 - hinge, 37, 278
 - Huber, 37
 - nonconvex, 276
 - quadratic, 34
 - ramp, 279
 - transduction, 287
- Marginal likelihood, 219
- Matching pursuit, 257, 264
- Matching statistic, 81
- Matrix-vector multiplication, 54, 178, 217, 231
 - sparse, 54
- Maxent, 64
- Maximum violating pair, 19, 22
- MILDE, 118
- MINOS, 27
- MKL; *see* Multiple kernel learning

- MNIST, 69, 131, 273, 298, 310, 319, 349, 350
- Modified finite newton, 157
- Modified gradient projection, 14, 15
- MPI, 125
- Multicast, 124
- Multiple kernel learning, 93
- Multipole methods; *see* Fast multipole methods
- Multithreading, 115
- MVM; *see* Matrix-vector multiplication

- N*-body learning, 26, 177, 201
- n*-mers, 75
- Network overhead, 122
- Newton method, 14, 34, 255
- No-free-lunch theorem, 324
- Nonconvex optimization, 275
- NORB, 352
- Numerical accuracy, 25

- Online learning, 303
- Optimal hyperplane, 3
- Optimality criteria, 8
- Optimization
 - convex, 6, 275
 - dual, 1, 105, 139
 - nonconvex, 275
 - primal, 29
- Parallel
 - cascade SVM, 109
 - chunking, 84
 - gradient projection, 109
 - kernel, 107, 114
 - scaling, 146
 - linear, 150
 - superlinear, 127
 - sequential learning, 144
 - spread kernel, 119, 121
- Perceptron, 53, 142, 333
 - on a budget, 306
- Preconditioning, 45
- Primal optimization, 29
 - advantages, 44
- Projection on the constraints, 14
- PVM, 125

- QP; *see* Quadratic programming
- Quadratic loss, 34
- Quadratic programming, 1, 106, 142

- Ramp loss, 279
- RCV1, 160, 296
- Reduced kernel expansion, 47, 251, 281
- Regression
 - probabilistic, 204
- Representer theorem, 33
- Reproducing kernel Hilbert space, 5, 228
- RKHS; *see* Reproducing kernel Hilbert space
- Royal Holloway SVM, 27

- Selective sampling, 301, 307
- Semi-infinite linear programming, 93
- Semi-supervised, 285
- Semisupervised, 161
- Sequence, 75
- Sequential minimal optimization, 18, 19, 53, 111
- SHOGUN, 86
- Shrinking, 23, 66, 115, 259
- SILP; *see* Semi-infinite linear programming
- SimpleSVM, 27
- SMO; *see* Sequential minimal optimization
- Soft margins, 5
- Sorted array, 79
- Sparse
 - dataset, 51, 157
 - expansion, 10, 251, 279, 282
 - feature maps, 79
- Spread kernel
 - full data, 119
 - split data, 121
- Stopping criterion, 20, 119, 167

- String kernel, 74
 - bag-of-words, 75
 - n -gram kernel, 75
 - spectrum, 76
 - weighted degree, 76
 - weighted degree with shifts, 77
- Suffix tree, 81
- Support vector machine, 301
 - complexity, 11, 280
 - dual objective function, 5, 7, 111
 - greedy optimization, 251
 - linear, 155, 157
 - nonconvex, 277, 279
 - online, 304
 - optimization problem, 5
 - parallel solver, 105, 139
 - primal objective function, 5, 32
 - primal optimization, 251
 - primal training, 41
 - sequential learning, 144
 - solver, 1, 29, 105, 139, 251
 - transductive; *see* Transductive SVM
- Support vectors, 10
 - bounded, 10
 - free, 10
 - number of, 11, 278
 - virtual, 319
- SVM; *see* Support vector machine
- SVM^{light}, 27, 86, 270, 292
- SVM^{lin}, 168
- SVQP, 27
- SVQP2, 27, 118

- Tangent vectors, 310
- Template matching, 335
- Transductive SVM
 - balancing constraint, 165, 288
 - CCCP, 288
 - CS³VM, 292
 - linear, 162
 - multiswitch, 161
 - ∇ T SVM, 291
 - nonconvex, 285
 - SVM^{light}, 291
- Tries, 80
- TSVM; *see* Transductive SVM

- Webb spam corpus, 86
- Working set method, 13, 16, 141
- Working set selection, 21
 - maximum gain, 21
 - maximum violating pair, 22
 - second-order schemes, 22