
Index

- alignment, 73
- all-subsequences kernel, 351
- all-subsets kernel, 294
- ANOVA kernels, 297

- backwards–forwards algorithm, 433
- bag-of-words, 328
- basic classification function, 120
- batch algorithm, 241
- batch ranking, 256
- binary classification, 20
- birthday paradox, 93
- box constraint algorithm, 224

- canonical correlation analysis, 164
- capacity, 93, 94
- centering data, 115
- centre of mass, 89
- character weightings string kernel, 368
- Cholesky decomposition, 58, 126
 - incomplete, 128
- Cholesky implementation, 126
- classification, 37, 211
- cluster matrices, 274
- cluster quality measure, 266
- clustering, 21, 39, 264
- compression, 244
- concentration, 88
- concentration inequalities, 86
- conditional independence, 400
- conjugacy, 163
- conjugate, 181
- conjugate gradient, 183
- conservative algorithm, 241
- consistent hypothesis, 212
- constraint
 - box, 221
- convolution kernel, 390
- corpus, 328
- correlation, 130
- covariance, 129
- covariance kernel, 66
- covariance matrix, 131

- data model, 399
- decomposition structure, 391
- deflation, 55, 123
- denoising, 146
- derived subsets kernel, 319
- determinant, 59
- dictionary, 328
- diffusion kernels, 310
- distance between feature vectors, 113
- distance from centre of mass, 114
- document–term matrix, 329
- dominating cost function, 98
- dot product, 48
- dual form of CCA, 168
- dual Lagrangian, 198
- dual partial least squares, 187
- dual solution, 32
- dual variables, 32
- duality, 32
- duality gap, 217
- dynamic programming, 300

- eigen-decomposition, 55
- eigenvalue, 54
- eigenvector, 54
- epsilon insensitive regression, 234
- expectation, 10
- expected distance from centre of mass, 114
- exponential diffusion kernel, 313

- feasibility gap, 227
- finitely positive semi-definite, 60
- Fisher discriminant, 132, 176
- Fisher information matrix, 424
- Fisher kernel, 425
- Fisher kernels, 421
- Fisher kernels for HMMs, 430
- Fisher score, 424
- fixed length Markov model Fisher kernel, 435
- Frobenius inner product, 73
- Frobenius norm, 73
- functional margin, 102

- Gaussian kernel, 77, 296
- generalisation error, 12, 86
- generalised eigenvector problem, 161
- geometric margin, 213
- Gram matrix, 32, 52
- Gram-Schmidt orthonormalisation
 - partial, 128
- Gram-Schmidt orthonormalisation, 124
- graph kernel, 305

- hard margin support vector machine, 213
- Heaviside function, 98
- hidden Markov model, 403
- hidden Markov model kernel
 - fixed length, 409
- hidden tree model kernel, 419
- hidden tree model kernels, 416
- Hilbert space, 48
- hinge loss, 99
- Hoeffding's inequality, 88
- hyperplane, 37
- hypersphere, 196

- insensitive loss, 234
- insensitive loss functions, 234
- intersection kernel, 315

- k-means algorithm, 273
- Karush-Kuhn-Tucker (KKT) conditions, 199
- kernel, 34
- kernel adatron
 - new, 247
- kernel canonical correlation analysis, 169
- kernel graph, 305
- kernel matrix, 52
- kernel methods, 17
- kernel normalisation, 112
- kernel PCA, 150
- kernel regularised CCA, 172
- kernels for trees, 383
- Kronecker delta, 51
- Krylov space, 183
- Kuhn-Tucker Theorem, 198

- latent semantic kernels, 338
- learning methodology, 4
- least squares approximation, 29
- leave one out, 246
- likelihood, 423
- linear ranking rule, 254
- linear time evaluation, 380
- linearly separable, 212
- Lipschitz, 98
- loss function, 9, 19

- margin, 102
- margin slack vector, 220
- marginalisation kernel, 401
- matric space, 49
- maximal margin hyperplane, 213
- maximising covariance, 157
 - maximising variance, 143
 - maximum covariance, 155
 - McDiarmid's theorem, 88
 - Mercer's theorem, 63
 - multiclass classification, 20
 - multidimensional scaling, 281

- neural networks, 37
- new support vector machine, 225
- norm of centre of mass, 113
- norm of feature vectors, 112
- normalised kernel, 40
- novelty detection, 21, 196
- nu-ranking, 259

- on-line classification, 241
- on-line learning, 241, 261
- on-line ranking algorithm
 - new, 262
- on-line regression, 241
- on-line support vector regression
 - new, 248
- operator
 - positive, 64
- orthonormal matrix, 56
- overfitting, 15
- overfitting CCA, 169

- P-kernel, 398
- p-spectrum, 347
- pair hidden Markov model, 412
- pair hidden Markov model kernel
 - fixed length, 415
- paired dataset, 165
- parametrised model class, 423
- partial least squares, 180
- pattern analysis algorithm, 12
- pattern function, 86
- perceptron, 37
- perceptron algorithm, 261
 - new, 242
- permutation testing, 87
- phylogenetic profiles, 419
- polynomial kernel, 292
- positive definite matrices, 57
- positive semi-definite matrices, 48, 57
- power method, 185
- primal solution, 32
- principal axes, 146
- principal components analysis, 38, 143
- principal components regression, 156, 177
- principal coordinates, 146
- projection, 123
- properties of the centre of mass, 114

- QR-decomposition, 125

- Rademacher, 92
- Rademacher complexity, 95
- Rademacher complexity of kernel classes, 99
- random variable, 87

- random vector, 87
- randomisation, 87
- rank, 52
- ranking, 20, 253
- Rayleigh quotient, 54
- RBF kernel, 77
- real number kernels, 317
- regression, 20, 230
 - algorithm, 2-norm, 232
- regularisation, 30, 93
- regularisation of CCA, 172
- regularised Fisher discriminant, 133
- regularised kernel Fisher discriminant, 136
- reproducing kernel Hilbert space, 63
- residual, 148
- ridge regression, 31
- robustness, 13

- scalar product, 48
- Schur product, 76
- semantic kernels, 333
- simple novelty detection, 116
- singular value decomposition, 141
- slack variable, 102, 203
- slack variables
 - variables, 220
- soft margin
 - algorithm, 1-norm, 221
 - algorithm, 2-norm, 227
 - optimisation, 219
- soft matching string kernel, 370
- soft minimal hypersphere, 203, 206
- soft ranking, 258
- solving CCA, 167
- sparseness, 198, 243
- spectral clustering, 274
- spectrum kernel, 348
- stability, 13
- stability analysis of CCA, 170
- stability of centering, 116
- stability of novelty detection, 119
- standardised data, 130
- statistically stable, 13
- stemming, 335

- stop words, 330
- string, 345
- string kernel, 360
- structured data, 382
- subsequence, 346
- subsequences kernel, 357
 - gap weighted, 361, 379
- subspace projection, 80
- substring, 345
- supervised learning, 19
- support vector machines, 211
- support vector regression
 - 1-norm, 238
 - 2-norm, 236
 - new, 240
- support vectors, 199, 216

- term, 328
- term–document matrix, 329
- test set, 17
- tf–idf representation, 334
- trace, 59
- training set, 22
- transduction, 20
- tree kernel
 - all-subtree, 387
 - co-rooted subtree, 385
- trie, 372
- trie-based kernels, 372

- uniform convergence, 94
- unitary matrix, 56
- unsupervised learning, 21

- variance, 130
- variance of projections, 132
- vector space kernel, 331
- vector space model, 328
- visualisation, 280
- von Neumann kernel, 313
- Voronoi diagram, 265

- weighting by number of gaps, 371
- whitening, 81