
Index

- χ^2 (critère du), 418
- θ -subsomption, 170, 172, 177
- θ -subsomption relative, 179

- absorbtion, 182
- abstraction, 211–213
- accuracy*, 120
- action de l’agent, 535
- actor-critic
 - apprentissage par renforcement, 557
- ADABOOST, 584
- AIC
 - critère de pénalisation, 702
- Akaike
 - critère d’information, 702
- algorithme *cluster and label* pour l’apprentissage semi-supervisé, 641
- algorithme AESA, 523
- algorithme Apriori, 620
- algorithme BRIG, 250
- algorithme d’élimination des candidats, 152
- algorithme d’apprentissage, 60
- algorithme d’auto-apprentissage, 642
- algorithme d’extension des étiquettes pour l’apprentissage semi-supervisé., 651
- algorithme de Baum-Welsh, 404
- algorithme de condensation, 521
- algorithme de nettoyage, 520
- algorithme de PLI générique, 185
- algorithme de propagation des étiquettes pour l’apprentissage semi-supervisé, 651
- algorithme de rétropropagation du gradient de l’erreur, 344
- algorithme de Viterbi, 401
- algorithme ECGI, 247
- algorithme EM, 402, 608, 638
 - pour l’apprentissage semi-supervisé, 640
- algorithme k-RI, 246
- algorithme RPNI, 251
- algorithme RPNI stochastique, 263
- algorithme *forward-backward*, 397

- algorithme PROGOL, 191
- algorithmes génétiques, 271, 273–284
 - espace génotypique, 271, 272, 274, 295
 - espace phénotypique, 272
 - fonction de performance, 278
 - opérateurs de croisement, 275, 277
 - opérateurs de mutation, 276, 277
 - pression sélective, 278
 - problèmes trompeurs, 284
 - remplacement, 280
 - sélection, 277–280
 - sélection proportionnelle à la performance, 278
 - sélection proportionnelle au rang, 279
 - sélection par tournoi, 279
 - systèmes de classeurs, 297
- algorithmes rapides pour les k-plus proches voisins, 523
- all-subsets selection*, 324
- alphabet auxiliaire d’une grammaire, 230
- alphabet terminal d’une grammaire, 230
- AM *Automated Mathematician*, 288
- analogie (raisonnement par), 213–214
- analyse en composantes principales communes, 99
- analyse en composantes principales, 99
- analyse en composantes indépendantes, 99
- analyse en composantes indépendantes, 625–628
- analyse en composantes principales, 624
- analyse en pire cas, 62
- analyse en variables latentes, 630
- ANOVA, 128
- apprentissage
 - à partir d’échecs, 208, 209
 - à partir d’explications (EBL), 204–211
 - batch*, 723
 - à grande échelle, 670
 - actif, 16, 723
 - auto-apprentissage, 636
 - avec bruit, 708

- co-apprentissage, 642–644
- collaboratif, 723
- dans les HMM, 402
- de classement, 98
- de sorties structurées, 680–683
- en ligne, 15, 60
- faible, 582, 590
- grande échelle, 717
- hors ligne, 15
- incrémental, 723
 - mistake-bound learning model*, 674
 - relative loss bound model*, 674
 - using expert advice*, 675
- multi-étiquette, 134
- multi-instance, 683
- multi-objectif, 133
- non supervisé, 17
- PAC, 62
 - de grammaires, 240–243
- par évolution simulée, 267–299
- par différences temporelles, 548–552
- par renforcement, 531–567
- par requête d'appartenance, 16
- par transfert, 672–673
- paresseux, 530
- petite échelle, 717
- semi-supervisé, 633–656
- supervisé, 14, 17, 41
 - de concepts, 22
- transductif, 637, 656
- apprentissage actif, 660–667
 - pool-based*, 660
 - stream-based*, 660
- apprentissage en-ligne, 673–675
- apprentissage hors-ligne (batch), 669
- apprentissage par renforcement
 - actor-critic, 557
 - bayésien, 557
 - hors politique, 551
 - sur politique, 550
- arbitrary binary queries*, voir protocole de requêtes binaires arbitraires
- arbres de décision, 105, 413–427, 529
- arbres de décision obliques, 421
- arbres de régression, 427–429
- astuce de la représentation unique, 145, 169
- attribut
 - arborescent, 140
 - binaire, 86, 140
 - catégoriel, 86
 - nominal, 86, 140
 - nominal hiérarchique, 86
 - nominal totalement ordonné, 87
 - numérique, 140
 - ordinal, 86
 - séquentiel nominal, 87
 - séquentiel numérique, 88
- auto-apprentissage, 636
- automates à états finis, 285
- bagging*, 590–591
- bandit à deux bras, 539
- bandits multi-bras, 558
- bandits (problème des), 539
- batch learning*, 669
- bayésienne (approche), 47, 50
- Bayes
 - classificateur naïf de, 493
 - formule de, 47, 50, 493
 - règle de, 45, 53
- Bellman
 - Principe d'optimalité, 537
- Bellman (équation de), 544
- biais, 67
- biais de recherche, 186
- biais de représentation, 26
- biais sémantique, 187
- biais syntaxique, 187
- BIC
 - Bayesian Information Criterion*, 703
 - critère de pénalisation, 703
- blending*, 214
- boosting*, 582–590
- bootstrap*, 118, 584, 679
- bruit de classe, 131
- bruit de mesure, 131
- cadre du regret, 673–675
- capacité (d'un espace d'hypothèses), 695
- cartes auto-organisatrices de Kohonen, 99, 601–602
- cascading*, 591
- causalité (apprentissage de la), 382
- chaînes de Markov, 107
- chaînes de Markov cachées, voir HMM
- chemin de régularisation, 327
- classificateur naïf de Bayes, 493
- classification supervisée, 41
- clause, 174
 - but, 175
 - définie, 174
 - de Horn, 174
 - unitaire, 175
- co-évolution, 295–297
- co-apprentissage, 642–644
- coûts non symétriques, 131
- codes correcteurs, 577–582
 - décodage par marges, 580
 - décodeur de Hamming, 580

- coefficient de pulvérisation, 696
cognitivism, 12–14
comité d'experts, 584
comparaison
 d'algorithmes, 125–129
 de méthodes d'apprentissage, 124
complexité algorithmique, voir complexité de Kolmogorov
complexité computationnelle dans l'apprentissage PAC, 722
complexité de Kolmogorov, 756–757
complexité de Rademacher, 697–698
compromis biais-variance, 67
confiance, voir règles d'association
confiance (règle d'association), 622
connexionnisme
 reservoir computing, 351–354
consistance universelle, 61
corrélation de Pearson, 685
corrélation de Spearman, 688
couverture d'un motif, 618
couverture de Markov, 95–96
couverture d'une hypothèse, 142
critère d'opérationnalité, 205
critère de performance, voir performance (mesure de)
critère inductif, 49–60
croisement, 277
cybernétique, 11–12
- d-séparation, 365–367, 372
décision optimale, 45
data streams, 112
data mining, 614–624
data mining, voir fouille de données
data warehouse, 614
deep belief networks, 349–351
déplacement covarié, 670–671
dérive de concept, 672
diagnostic, 371
dilemme exploitation vs. exploration, 276, 539–541
dimension
 de Vapnik-Chervonenkis, 697
 réduction de, 89–99
dimension de Vapnik-Chervonenkis, 747, 748
distance, 597
distributions conjuguées, 54
données bruitées, 131
données déséquilibrées, 129–131
données en grande dimension, 131–132
données non indépendantes, 132–133
données temporelles, 353
- early stopping rule*, 708
- EBL (*Explanation-Based Learning*), 204–211
échantillon
 d'apprentissage, 22, 121
 de test, 121
 de validation, 121
effet tunnel cognitif, 214
élagage d'un arbre de décision, 422
entrepôt de données, 614
entropie (calcul sur les flux), 677
entropie croisée, 95, 498
épistasie, 284
equivalence query, voir protocole de requêtes d'équivalence
erreur
 d'approximation, 67, 717
 d'estimation, 67
 d'optimisation, 717
 de biais, 717
 de première espèce, 116
 de seconde espèce, 116
espace de Hilbert à noyau reproduisant, 443
espace de redescription, 437
espace de représentation, 85–88
espace des versions, 150
espace génotypique, 271, 272, 274, 295
espace phénotypique, 272
estimation de densité, 59
évaluation des performances
 courbe ROC, 121
 par échantillon de test, 115
 par la méthode du *leave-one-out*, 118
 par resubstitution, 115
 par validation croisée, 117
 par *bootstrap*, 118
 par *jackknife*, 118
évolution simulée, 271
exemple, 8
explication, 371
exploitation, voir dilemme exploitation vs. exploration, 539
exploration, voir dilemme exploitation vs. exploration, 540
extraction de connaissances des données, xi
- F_{mesure}, 120
factorisation matricielle non-négative (NMF), 628
fat-shattering dimension, 700
fenêtres de Parzen, 512
filtrage collaboratif, 684–690
Fisher
 information de, 662
fitness function, 272, 278
flot de données, 112
flux de données

- échantillonnage, 676
- clustering*, 679
- min-wise sampling*, 677
- reservoir sampling*, 676
- calcul d'entropie, 677
- détection de rupture, 678
- résumés, 677
- flux de données (apprentissage à partir de), 675–677
- FOIL, 188–190
- fonction cible, 20
- fonction de croissance, 747
- fonction de performance, 278
- fonction de perte, 44
 - coude, 452
 - de substitution, 452
- fonction de risque, 15
- fonction de score, 378
- fonction noyau, 512
 - apprentissage de, 483
- fonction sigmoïde, 333
- fonctions séparatrices, 104
- fouille de données, xi, 132, 614–624
- fréquence d'un motif, 618

- généralisé maximale spécifiquement, 146
- généralité (relation de), 142
- générative (approche), 47
- Gauss-Markov (théorème de), 309
- Gini (critère de), 418
- gradient stochastique, 669–670
- grammaire formelle, 230
- grammaire hors-contexte, 231
- grammaire k-réversible, 246
- grammaire régulière, 231
- grande dimension (phénomènes), 90
- graphe
 - laplacien de, 476, 651
- G-set*, 150

- Hadamard Jacques, 18
- heuristiques de contrôle, 204, 209
- hinge loss*, 452
- HMM, 107, 387–409
 - apprentissage des paramètres d'un, 402
 - ergodique, 395
 - gauche-droite, 395
 - probabilité d'observation, 394
 - probabilité de transition, 394
 - probabilités initiales, 394
- Hoëfdding
 - inégalité de, 64
- Hopfield
 - réseaux de, 351
- hors politique
 - apprentissage par renforcement, 551
- hypothèse
 - cohérente, 144
 - complète, 142, 144
 - correcte, 142, 144
- hypothèse nulle, 116
- hypothèse optimale, 45

- i.i.d. (tirage), 19
- identification, 182
- inductive logic programming (ILP)*, voir programmation logique inductive (PLI)
- inférence grammaticale, 223–264
- information de Fisher, 662
- information mutuelle, 378, 627
- instance, 85
- intelligibilité, 124
- inter-construction, 182
- intervalle de confiance, 116
- intra-construction, 182
- invention de prédicat, 182
- inversion de la résolution, 181–184
- itemsets, 132
- itemsets fréquents, 111

- jackknife*, 118
- jeu de Go, 563–564

- k-plus-proches-voisins (règle des), 516
- Karush-Kuhn-Tucker (conditions de), 449
- KDD (*knowledge discovery in databases*), voir extraction de connaissances des données
- kernel methods, 324
- kernel trick, 438
- Kullback-Leibler (divergence de), 95, 678, 702
- kurtosis, 626

- lagrangien, 448
- langage engendré par une grammaire, 230
- langage hors-contexte, 231
- langage régulier, 231
- Langley Pat, 19
- laplacien
 - d'un graphe, 477
- large-scale learning*, 717
- large-scale learning*, 670
- learning to rank*, 98
- leave-one-out*, 118
- Levi-Strauss Claude, 12
- LEX, 155–156
- linéaire par morceaux (modèle), 321
- log-vraisemblance, 51
- logique de description, 110
- logique des prédicats, 110
- logique des propositions, 109

- loi des grands nombres, 63
- mélange de gaussiennes, 739
- mémoire associative, 352
- méthodes gloutonnes, 550
- méthodes semi-paramétriques, 525
 - discrimination logistique, 525
 - mélange de distributions, 526
- macro-opérateurs, 204, 209
- MAE (mean average error), 689
- marge, 580
- marge d'un exemple, 587
- marge douce, 452
- marge géométrique, 447
- Markov
 - équivalence de, 368
 - condition de, 363
 - couverture de, 368
 - processus décisionnels de, 538
- Markov blanket*, 95–96
- matrice de codage des classes, 577
- matrice de confusion, 115
- matrice de Gram, 445
- matrice positive semidéfinie, 439
- maximum a posteriori (MAP), 55
- maximum de vraisemblance, 50, 55
- Maximum Likelihood Estimator (MLE), 55
- MDLP (*Minimum Description Length Principle*), 52, 379, 704, voir principe de longueur minimale de description
- membership query*, voir protocole de requêtes d'appartenance
- mesure de dissimilarité, 597
- méta-apprentissage, 573
- méta-paramètre, 573
- méthodes non paramétriques, 497, 510
- méthodes spectrales, 610–611
- min-wise sampling*, 677
- minimisation du risque empirique, 49, 50
- minimisation du risque structurel (*SRM*), 711
- minimisation du risque structurel (*SRM*), 710
- mixture models, 323
- modèle
 - sélection automatique, 700
- modèle de mélange, 323
- modèle de Markov caché, voir HMM
- modèles graphiques, 107
- moindres carrés (estimation des), 306
- Monte Carlo (méthode de), 547
- mot engendré par une grammaire, 230
- motif, 618
 - couverture de, 618
 - fréquence de, 618
 - support de, 618
- motif fréquent, 616
- motifs fréquents, 617–621
- mots dérivés selon une grammaire, 230
- multiplicateurs de Lagrange, 448, 707–708
- mutation, 276, 277
- négentropie, 626
- neurone formel, 332
- no-free-lunch theorem*, 74–277
- nombre de couverture, 698–700
- non-negative matrix factorization (NMF)*, 628
- non-paramétriques (méthodes), 497, 510
- noyau
 - ANOVA, 465
 - arbres, 472
 - astuce des, 438
 - de convolution, 471
 - de Fisher, 479
 - diffusion, 476
 - génératif, 478–479
 - gaussien, 466
 - graphes, 473
 - reproduisant, 443
 - séquences, 470
 - spectre, 469
 - sur des ensembles, 467
 - sur des séquences, 468
 - sur des textes, 468
 - toutes sous-séquences, 471
 - universel, 461
- noyaux (méthodes à), 324
- observation, 14
- Occam (algorithme d'), 760–761
- Occam (rasoir d'), 704
- off policy
 - apprentissage par renforcement, 551
- on policy
 - apprentissage par renforcement, 550
- one-class SVM*, 459
- ontologie, 107
- opérateurs de généralisation, 147–149
- opérateurs de spécialisation, 148
- optimisation de performance, 17
- oracle, 14
- parcimonieux (codage), 631
- Pearson (corrélation de), 685
- performance (mesure de), 272
- pertinence de l'*ERM*, 65
- physique statistique (analyse de la), 718–721
- plan d'expériences, 666
- pointwise estimation*, 55
- politique, 531, 535
- politique optimale, 542

- POMDP (Partially Observable MDP), 561
pool-based active learning, 660
 poursuite de projections, 628
 précision, 120
 prédiction, 15
 prédiction de séquence, 285
 prétraitement des données, 88
 présentation
 complète, 238
 négative, 238
 positive, 238
 pression sélective, 278
 principe de longueur minimale de description, 757
 principe de longueur minimale de description, 52, 379, 499, 704, 759
 principe de minimisation du risque structurel, 756
 problèmes trompeurs, 284
 processus décisionnels de Markov, 538
 processus de Markov, 390
 PRODIGY, 210
 PROGOL, 191–192
 programmation évolutive, 271
 programmation génétique, 271, 285–294
 programmation logique inductive (PLI), 163–199
 proie-prédateur, 296
projection pursuit, 628
 protocole d'apprentissage, 15
 protocole de requêtes d'appartenance, 661
 protocole de requêtes binaires arbitraires, 661
 protocole de requêtes d'équivalence, 661
 protocole de requêtes statistiques, 661
 pseudo-dimension, 700
 puissance du test, 116

Q-learning, 551
query-by-committee, 666

 régularisation
 exemple, 424
 théorie de la, 706–708
 réseaux bayésiens, 95, 107, 358–384
 équivalence de Markov, 368
 apprentissage des, 373–381
 condition de Markov, 363
 couverture de Markov, 368
 d-séparation, 365–367, 372
 diagnostic, 371
 explication, 371
 fonction de score, 378
 méthode des arbres de jonction, 372
 méthode EM, 376
 méthodes de conditionnement, 372
 méthodes de Monte Carlo, 372
 raisonnement dans les, 369–372
 réseaux connexionnistes, 329–355
 apprentissage, 338–349
 architecture multicouche, 334
 choix de l'architecture, 349
 exemple de fonctionnement, 335
 réseaux de neurones
 profonds, 349–351
 résumés (calcul sur les flux), 677
 rétropropagation du gradient de l'erreur, 344
 régression, 58–59, 305–306
 aux moindres carrés, 497
 de but, 206
 régression linéaire, 305–310
 règle de production d'une grammaire, 230
 règles d'association, 111–112, 616, 621–623, 677–678
 confiance, 678
 support, 677
 règles de contrôle, 204
 raisonnement à partir de cas (RàPC), 214–216
ranking, 59, 98
 rappel, 120
recall, 120
 reconnaissance des formes, xi
 redescription, 91
 espace de, 437
 reformulation, 212
 règles d'association
 fortes, 622
 remplacement, 280
 renforcement (signal de), 534, 535
 représentation duale, 436
 représentation par attribut-valeur, 110
representer theorem, 444
reservoir sampling, 676
reservoir computing, 351–354
 resubstitution (estimation de l'erreur), 115
ridge regression, 324, 325
 risque empirique, 50
 risque réel, 45
 RKHS, voir espace de Hilbert à noyau reproduisant
 robot simulé, 292–294
 ROC (courbe), 121

 sélection, 277–280
 sélection proportionnelle à la performance, 278
 sélection proportionnelle au rang, 279
 sélection (étape de), 278
 sélection automatique de modèle, 700
 sélection d'attributs, 90
 sélection de modèle, 52
 sélection par tournoi, 279

- séries temporelles, 353
- sac de mots, 467
- SARSA, 550
- sciences de l'artificiel, 5
- Shannon Claude, 11
- simplicité, 124
- single representation trick*, 145, 169
- SLD-résolution, 176
- SLD-subsumption, 181
- small-scale learning*, 717
- SOAR, 209
- soft margin*, 452
- Solomonoff (induction selon), 756
- sortie désirée, 14
- spécialisé maximale général, 146
- sparse coding*, 631
- Spearman (corrélation de), 688
- speed-up learning*, 208
- spline (régression), 322
- S-set*, 150
- stacked generalization*, 576–577
- stepwise selection*, 324
- stratégies d'évolution, 271, 285
- stream-based active learning*, 660
- subgroup discovery*, 624
- subsumption (relation de), 170
- support, voir règles d'association
- support (règle d'association), 622
- support d'un motif, 618
- sur politique
 - apprentissage par renforcement, 550
- sur-adaptation, 9
- sur-apprentissage, 9
- sur-complète (représentation), 631
- surrogate functions*, 452
- SVM
 - exemples critiques, 449
 - lagrangien, 448
 - mono-classe, 459
 - multiplicateurs de Lagrange, 448
 - théorème de Mercer, 440
 - variables duales, 448
- SVM transductifs (S3VMs), 645–647
- symétrisation (astuce de la), 698
- système apprenant, 14
- systèmes de classeurs, 297
- systèmes de recommandation, 684
- systèmes dynamiques, 721

- TD(λ), 551–552
- Test
 - de Friedman, 129
- test
 - de Wilcoxon, 127
 - multiple, 65
 - statistique
 - score Z , 308
- tf-idf, 468, 688
- théorème central limite, 64
- théorème de Mercer, 440
- théorème de représentativité, 444
- théorème des schémas, 280–284
- trace d'éligibilité, 551
- transition de phase, 195–197
- treillis, 146
- tri, 59
- Turing Alan, 11

- UCB, 558
- UCT, 558–561
- utilité
 - fonction d', 538, 541
 - problème de l', 210

- validation croisée, 117–118
- Vapnik-Chervonenkis (dimension de), 697
- variables duales, 448
- variables latentes, 90, 381
- variance, 67
- vecteurs de support, voir exemples critiques
- vector space model*, 467
- vote, 576–577
- vraisemblance, 51
 - fonction de, 374

- weight decay*, 708
- Wiener Norbert, 11
- Wilcoxon signed rank test*, 127