

L'APPRENTISSAGE ARTIFICIEL

CONCEPTS ET ALGORITHMES

(2ÈME ÉDITION)

Antoine CORNUÉJOLS

et

Laurent MICLET

Editions Eyrolles, 2010, 55 €

Table des matières

Table des matières	vii
Avant-propos	xiii
Quelques applications de l'apprentissage artificiel	xiv
Quelques définitions de base	xiv
Deux champs industriels de l'apprentissage artificiels : la reconnaissance des formes et la fouille de données	xv
Les caractéristiques de l'apprentissage artificiel	xvii
Trois exemples d'apprentissage	xviii
Organisation et plan de l'ouvrage	xxi
Guide de lecture	xxiii
Notations	xxvii
I Les fondements de l'apprentissage	1
1 De l'apprentissage naturel à l'apprentissage artificiel	3
1 L'apprentissage artificiel	4
2 Deux exemples : apprendre à jouer, apprendre à lire	6
3 Deux approches : la cybernétique et les sciences cognitives	10
4 Les concepts de base de l'apprentissage	14
5 L'induction comme un jeu entre espaces	22
6 Retour sur l'organisation de l'ouvrage	29
2 Première approche théorique de l'induction	37
1 Poser un problème d'apprentissage	40
2 Approches bayésiennes et approche directe pour décider	46
3 Le critère inductif	49
4 Analyse du principe de <i>minimisation du risque empirique</i>	60
5 Le lien entre le passé et le futur et le <i>no-free-lunch theorem</i>	73
6 Notes historiques et bibliographiques	78

3	L'environnement méthodologique de l'apprentissage	81
1	L'espace des données d'apprentissage	85
2	L'espace des hypothèses d'apprentissage	101
3	Les protocoles d'apprentissage	112
4	L'évaluation de l'apprentissage	113
5	La comparaison des méthodes d'apprentissage	124
6	Autres problèmes pratiques	129
II	Apprentissage par exploration	135
4	Induction et relation d'ordre : l'espace des versions	137
1	Les concepts de base	140
2	La structuration de l'espace des hypothèses	144
3	La construction de l'espace des versions	152
4	La représentation des connaissances par un treillis de Galois	156
5	La programmation logique inductive	161
1	La programmation logique inductive : le cadre général	165
2	La logique des prédicats et les programmes logiques : terminologie	171
3	La structuration de l'espace des hypothèses en logique des prédicats	175
4	L'exploration de l'espace des hypothèses	182
5	Deux exemples de systèmes de PLI	186
6	La probabilisation de la PLI	190
7	Les domaines d'application de la PLI	191
8	Les chantiers de la PLI	193
6	Transfert de connaissances et apprentissage par analogie	199
1	L'apprentissage en présence de théorie	200
2	L'apprentissage par examen de preuve (<i>EBL</i>)	201
3	Abstraction et reformulation des connaissances	208
4	Changement de repère, raisonnement par analogie et RàPC	210
5	L'apprentissage par proportion analogique	213
6	Bilan	217
7	L'inférence grammaticale	219
1	Définitions et notations	225
2	Les protocoles de l'inférence : quelques résultats théoriques	233
3	L'espace de recherche de l'inférence régulière	240
4	L'inférence régulière sans échantillon négatif	241
5	L'inférence régulière sous contrôle d'un échantillon négatif	246
6	L'inférence de grammaires algébriques	250
7	L'inférence d'automates probabilistes	257
8	Quelques approches complémentaires	260

8	Apprentissage par évolution simulée	263
1	Trois espaces au lieu de deux	265
2	Un modèle formel simplifié de l'évolution	268
3	Les algorithmes génétiques	269
4	Les stratégies d'évolution	281
5	La programmation génétique	281
6	La coévolution	291
III	Apprentissage par optimisation	297
9	L'apprentissage de modèles linéaires	299
1	Introduction	300
2	Régression linéaire	301
3	Séparatrices linéaires	306
4	Modèles linéaires par morceaux et combinaisons de modèles locaux	317
5	La recherche des facteurs pertinents	320
10	L'apprentissage de réseaux connexionnistes	325
1	Les différents éléments d'un réseau connexionniste	328
2	L'architecture multicouche	330
3	L'algorithme d'apprentissage	335
4	Quelques résultats théoriques sur les réseaux connexionnistes	344
5	Comment choisir l'architecture d'un réseau ?	345
6	Les réseaux à architecture profonde	346
7	Réseaux et régime dynamique : le <i>Reservoir Computing</i>	348
11	L'apprentissage de réseaux bayésiens	353
1	Les modèles graphiques	355
2	Les réseaux d'inférence bayésiens	357
3	Les inférences dans les réseaux bayésiens	365
4	L'apprentissage des réseaux bayésiens	369
5	L'inférence de relations causales	378
6	Applications	379
7	Quelques logiciels	380
12	L'apprentissage de modèles de Markov cachés	383
1	Les modèles de Markov observables	386
2	Les modèles de Markov cachés (HMM)	387
3	Les HMM comme règles de classification de séquences	392
4	L'évaluation de la probabilité d'observation	393
5	Le calcul du chemin optimal : l'algorithme de Viterbi	395
6	L'apprentissage	398
7	Approfondissements	403
8	Applications	404
13	Apprentissage par inférence d'arbres	407
1	Les arbres de décision	409
2	Les arbres de régression	423

IV	Apprentissage par approximation et interpolation	427
14	Méthodes à noyaux	429
1	Trois voies vers les méthodes à noyau	431
2	Philosophie des méthodes à noyaux	441
3	Les Séparatrices à Vastes Marges (SVM)	443
4	Autres types d'induction avec fonctions noyau	453
5	Ingénierie des fonctions noyau	458
6	Les méthodes à noyaux en pratique	477
7	Bilan et perspectives	483
15	L'apprentissage bayésien et son approximation	487
1	L'apprentissage bayésien	490
2	Les méthodes paramétriques	499
3	Les méthodes non paramétriques	509
4	Les méthodes semi-paramétriques	523
16	L'apprentissage de réflexes par renforcement	531
1	Description du problème	534
2	Si tout est connu : l'utilité de la fonction d'utilité	541
3	L'apprentissage des fonctions d'utilité quand l'environnement est connu	543
4	Sans modèle du monde : la méthode de Monte-Carlo	547
5	La méthode des différences temporelles	548
6	La généralisation dans l'apprentissage par renforcement	552
7	Contrôle optimal par recherche arborescente et algorithme UCT	558
8	Le cas des environnements partiellement observables	561
9	Exemples d'application	562
10	Bilan et perspectives	565
V	Au-delà de l'apprentissage supervisé	569
17	Apprentissage de combinaisons d'experts	571
1	Principes des méthodes par combinaison	572
2	Le vote de plusieurs classificateurs	576
3	Les codes correcteurs de classes	577
4	Le <i>boosting</i> d'un algorithme d'apprentissage	582
5	Le <i>bagging</i>	590
6	Les forêts aléatoires (<i>random forests</i>)	591
7	L'apprentissage en cascade (<i>cascading</i>)	591
18	La classification non supervisée et la fouille de données	593
1	Introduction	596
2	Les méthodes de classification fondées sur les distances	597
3	Les méthodes de classification par des modèles probabilistes	609
4	Méthodes spectrales de catégorisation	610
5	La classification de données symboliques	612
6	La fouille de données	615
7	Les analyses en composantes	625

19	L'apprentissage semi-supervisé	635
1	Présentation	636
2	Les modèles génératifs : apprentissage dans l'espace joint $\mathcal{X} \times \mathcal{Y}$	640
3	L'auto-apprentissage	643
4	Le co-apprentissage	644
5	L'utilisation d'hypothèses fondamentales sur les données	646
6	Quelques directions pour l'apprentissage semi-supervisé	655
7	Conclusion et petite discussion	657
20	Vers de nouvelles tâches et de nouvelles questions	659
1	Apprentissage actif	661
2	Apprentissages en ligne, incrémental et par transferts	669
3	Apprentissage à partir de flux de données	677
4	Apprentissage de sorties structurées	682
5	Apprentissage pour le filtrage collaboratif	686
21	Analyse de l'induction : approfondissements et ouvertures	695
1	Généralisation de l'analyse du principe MRE	696
2	Principes inductifs contrôlant l'espace des hypothèses	702
3	Prise en compte de l'algorithme d'apprentissage dans la théorie	715
4	Autres types d'analyses	720
VI	Annexes techniques	727
22	Annexes techniques	729
1	Exemples de fonctions de perte en induction	729
2	Le calcul de l'intervalle de confiance de l'estimation de la probabilité d'une règle de classification	732
3	Estimation d'une densité de probabilité en un point.	733
4	L'estimation des paramètres d'une distribution gaussienne.	734
5	Pourquoi et comment la règle du PPV converge-t-elle?	735
6	Pourquoi la règle de décision bayésienne est-elle optimale?	736
7	Apprentissage par estimation-maximisation	736
8	Optimisation par descente de gradient	740
9	La rétropropagation du gradient de l'erreur	744
10	L'analyse de l'induction de Vapnik	747
11	L'induction par compression d'information	758
	Index	795

Index

- χ^2 (critère du), 414
- θ -subsumption, 168, 170, 175
- θ -subsumption relative, 177

- absorbtion, 180
- abstraction, 208–210
- accuracy*, 119
- action de l’agent, 535
- actor-critic
 - apprentissage par renforcement, 557
- ADABOOST, 584
- AIC
 - critère de pénalisation, 704
- Akaike
 - critère d’information, 704
- algorithme *cluster and label* pour l’apprentissage semi-supervisé, 643
- algorithme AESA, 522
- algorithme Apriori, 621
- algorithme BRIG, 247
- algorithme d’élimination des candidats, 152
- algorithme d’apprentissage, 61
- algorithme d’auto-apprentissage, 644
- algorithme d’extension des étiquettes pour l’apprentissage semi-supervisé, 653
- algorithme de Baum-Welsh, 400
- algorithme de condensation, 519
- algorithme de nettoyage, 519
- algorithme de PLI générique, 183
- algorithme de propagation des étiquettes pour l’apprentissage semi-supervisé, 652
- algorithme de rétropropagation du gradient de l’erreur, 340
- algorithme de Viterbi, 397
- algorithme ECGI, 243
- algorithme EM, 398, 610, 640
 - pour l’apprentissage semi-supervisé, 642
- algorithme k-RI, 243
- algorithme RPNI, 247
- algorithme RPNI stochastique, 259
- algorithme *forward-backward*, 393

- algorithme PROGOL, 189
- algorithmes génétiques, 267, 269–280
 - élection proportionnelle à la performance, 274
 - espace génotypique, 266, 268, 270, 291
 - espace phénotypique, 268
 - fonction de performance, 274
 - opérateurs de croisement, 271, 273
 - opérateurs de mutation, 272, 273
 - pression sélective, 274
 - problèmes trompeurs, 280
 - remplacement, 276
 - sélection, 273–276
 - sélection proportionnelle au rang, 275
 - sélection par tournoi, 275
 - systèmes de classeurs, 293
- algorithmes rapides pour les k-plus proches voisins, 522
- all-subsets selection*, 320
- alphabet auxiliaire d’une grammaire, 226
- alphabet terminal d’une grammaire, 226
- AM *Automated Mathematician*, 284
- analogie (raisonnement par), 210–211
- analyse en composantes principales communes, 99
- analyse discriminante linéaire, 506
- analyse discriminante non linéaire, 506
- analyse en composantes indépendantes, 99, 626–630
- analyse en composantes principales, 99, 626
- analyse en pire cas, 63
- analyse en variables latentes, 632
- ANOVA, 128
- apprentissage
 - à partir d’échecs, 205, 206
 - à partir d’explications (EBL), 201–208
 - à grande échelle, 672
 - actif, 16, 725
 - auto-apprentissage, 638
 - avec bruit, 710

- batch*, 671, 725
- co-apprentissage, 644–646
- collaboratif, 725
- dans les HMM, 398
- de classement, 98
- de sorties structurées, 456–458, 682–685
- en ligne, 15, 60
- faible, 582, 590
- grande échelle, 719
- hors ligne, 15
- incrémental, 725
 - mistake-bound learning model*, 676
 - relative loss bound model*, 676
 - using expert advice*, 677
- multi-étiquette, 133
- multi-instance, 685
- multi-objectif, 133
- non supervisé, 17
- PAC, 62
 - de grammaires, 236–239
 - par évolution simulée, 263–295
 - par différences temporelles, 548–552
 - par renforcement, 531–567
 - par requête d'appartenance, 16
 - par transfert, 674–675
 - paresseux, 529
 - petite échelle, 719
 - semi-supervisé, 635–658
 - supervisé, 14, 17, 41
 - de concepts, 22
 - transductif, 639, 657, 658
- apprentissage actif, 662–669
 - pool-based*, 662
 - stream-based*, 662
- apprentissage en ligne, 675–677
- apprentissage hors ligne (batch), 671
- apprentissage hors-ligne (batch), 725
- 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, 409–423, 528
- arbres de décision obliques, 417
- arbres de régression, 423–425
- astuce de la représentation unique, 145, 167
- attribut
 - arborescent, 140
 - binaire, 86, 140
 - catégoriel, 86
 - nominal, 86, 140
 - nominal hiérarchique, 87
 - nominal totalement ordonné, 87
 - numérique, 140
 - ordinal, 86
 - séquentiel nominal, 87
 - séquentiel numérique, 88
- auto-apprentissage, 638
- automates à états finis, 281
- bagging*, 590
- bandit à deux bras, 539
- bandit multi-bras, 558
- bandits (problème des), 539
- batch learning*, 671, 725
- bayésienne (approche), 46, 50
- Bayes
 - classificateur naïf de, 491
 - formule de, 47, 50, 491
 - règle de, 45, 53
- Bellman
 - Principe d'optimalité, 537
- Bellman (équation de), 544
- biais, 67
- biais de recherche, 184
- biais de représentation, 26
- biais sémantique, 185
- biais syntaxique, 185
- BIC
 - Bayesian Information Criterion*, 705
 - critère de pénalisation, 705
- blending*, 211
- boosting*, 582–590
- bootstrap*, 118, 585, 681
- bruit de classe, 131
- bruit de mesure, 131
- cadre du regret, 675–677
- capacité (d'un espace d'hypothèses), 697
- cartes auto-organisatrices de Kohonen, 99, 601–603
- cascading*, 591–592
- causalité (apprentissage de la), 378
- chaînes de Markov, 107
- chaînes de Markov cachées, voir HMM
- chemin de régularisation, 323
- classificateur naïf de Bayes, 491
- classification supervisée, 41
- clause, 172
 - but, 173
 - définie, 172
 - de Horn, 172
 - unitaire, 173
- co-évolution, 291–293
- co-apprentissage, 644–646
- coûts non symétriques, 131
- codes correcteurs, 577–582

- décodage par marges, 580
- décodeur de Hamming, 580
- coefficient de pulvérisation, 698
- cognitivism, 12–14
- comité d'experts, 583
- comparaison
 - d'algorithmes, 125–129
 - de méthodes d'apprentissage, 124
- complexité algorithmique, voir complexité de Kolmogorov
- complexité calculatoire dans l'apprentissage PAC, 724
- complexité de Kolmogorov, 758–759
- complexité de Rademacher, 699–700
- compromis biais-variance, 68
- confiance, voir règles d'association
- confiance (règle d'association), 623
- connexionnisme
 - reservoir computing*, 348–351
- consistance universelle, 61
- corrélation de Pearson, 687
- corrélation de Spearman, 690
- couverture d'un motif, 619
- couverture de Markov, 95–96
- couverture d'une hypothèse, 142
- covariate shift*, 672–673
- critère d'opérationnalité, 202
- critère de performance, voir performance (mesure de)
- critère inductif, 49–60
- croisement, 273
- cybernétique, 11–12

- d-séparation, 361–363, 368
- décision optimale, 45
- data streams, 112
- data mining*, voir fouille de données
- data warehouse*, 615
- deep belief networks*, 346–347
- déplacement covarié, 672–673
- dérive de concept, 60, 674
- diagnostic, 367
- dilemme exploitation vs. exploration, 272, 539–541
- dimension
 - de Vapnik-Chervonenkis, 699, 749, 751
 - réduction de, 89–99
- dimensionnalité
 - fléau de, 89, 505, 608
- discriminante (approche), 47–49, 493
- distance, 597
- distributions conjuguées, 54
- données bruitées, 131
- données déséquilibrées, 129–130
- données en grande dimension, 131
- données non indépendantes, 132–133
- données temporelles, 349

- early stopping rule*, 710
- EBL (*Explanation-Based Learning*), 201–208
- échantillon
 - d'apprentissage, 22, 120
 - de test, 120
 - de validation, 120
- effet tunnel cognitif, 211
- élagage d'un arbre de décision, 418
- EM (*Expectation - Maximization*), 736–740
- entrepôt de données, 615
- entropie (calcul sur les flux), 679
- entropie croisée, 95, 496
- épistasie, 280
- equivalence query*, voir protocole de requêtes d'équivalence
- erreur
 - d'approximation, 67, 719
 - d'estimation, 67
 - d'optimisation, 719
 - de biais, 719
 - de première espèce, 116
 - de seconde espèce, 116
- espace de Hilbert, 437, 438
- espace de Hilbert à noyau reproduisant, 439
- espace de redescription, 433
- espace de représentation, 85–88
- espace des versions, 150
- espace génotypique, 266, 268, 270, 291
- espace phénotypique, 268
- 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, 114
 - par validation croisée, 116
 - par *bootstrap*, 118
 - par *jackknife*, 118
- évolution simulée, 267
- exemple, 8
- explication, 367
- exploitation, voir dilemme exploitation vs. exploration, 539
- exploration, voir dilemme exploitation vs. exploration, 540
- extraction de connaissances des données, xv

- F_mesure, 120
- factorisation matricielle non-négative (NMF), 630
- fat-shattering dimension*, 702
- fenêtres de Parzen, 512

- filtrage collaboratif, 686–692
- Fisher
 - information de, 664
- fitness function*, 268, 274
- fléau de la dimensionnalité, 89, 505, 608
- flot de données, 112
- flux de données
 - échantillonnage, 678
 - clustering*, 681
 - min-wise sampling*, 679
 - reservoir sampling*, 678
 - calcul d'entropie, 679
 - détection de rupture, 680
 - résumés, 679
- flux de données (apprentissage à partir de), 677–679
- FOIL, 186–188
- fonction cible, 20
- fonction de croissance, 749
- fonction de performance, 274
- fonction de perte, 44
 - coude, 448
 - de substitution, 448
- fonction de risque, 14
- fonction de score, 374
- fonction noyau, 511
 - apprentissage de, 481
- fonction sigmoïde, 329
- fonctions séparatrices, 103
- fouille de données, xv, 132, 615–625
- fréquence d'un motif, 619

- généralisé maximalement spécifique, 146
- généralité (relation de), 142
- générative (approche), 47, 493
- Gauss-Markov (théorème de), 305
- Gini (critère de), 414
- gradient stochastique, 671–672
- grammaire formelle, 226
- grammaire hors-contexte, 227
- grammaire k-réversible, 242
- grammaire régulière, 227
- grande dimension (phénomènes), 90, 608
- graphe
 - laplacien de, 474, 653
- G-set*, 150

- Hadamard Jacques, 18
- heuristiques de contrôle, 201, 206
- Hilbert
 - espace de, 437, 438
- hinge loss*, 448
- HMM, 107, 383–405
 - apprentissage des paramètres d'un, 398
 - ergodique, 391
 - gauche-droite, 391
 - probabilité d'observation, 390
 - probabilité de transition, 390
 - probabilités initiales, 390
- Hoeffding
 - inégalité de, 64
- Hopfield
 - réseaux de, 348
- hors politique
 - apprentissage par renforcement, 551
- hypothèse
 - cohérente, 144
 - complète, 142, 144
 - correcte, 142, 144
- hypothèse nulle, 115
- hypothèse optimale, 45

- i.i.d. (tirage), 19
- identification, 180
- inductive logic programming (ILP)*, voir programmation logique inductive (PLI)
- inférence grammaticale, 219–261
- information de Fisher, 664
- information mutuelle, 374, 628
- instance, 85
- intelligibilité, 124
- inter-construction, 180
- intervalle de confiance, 116
- intra-construction, 180
- invention de prédicat, 180
- inversion de la résolution, 179–182
- itemsets, 132
- itemsets fréquents, 111

- jackknife*, 118
- jeu de Go, 563–564
- Johnson-Lindenstrauss
 - théorème de, 609

- k-plus-proches-voisins (règle des), 511, 515
- Karush-Kuhn-Tucker (conditions de), 445
- KDD (*knowledge discovery in databases*), voir extraction de connaissances des données
- kernel methods*, 320, 429–486
- kernel trick, 434
- Kullback-Leibler (divergence de), 95, 680, 704
- kurtosis, 627

- lagrangien, 444
- langage engendré par une grammaire, 226
- langage hors-contexte, 227
- langage régulier, 227
- Langley Pat, 19
- laplacien
 - d'un graphe, 475

- large-scale learning*, 719
- large-scale learning*, 672
- learning to rank*, 98
- leave-one-out*, 118
- Levi-Strauss Claude, 12
- LEX, 155–156
- linéaire par morceaux (modèle), 317
- log-vraisemblance, 51
- logique de description, 110
- logique des prédicats, 110
- logique des propositions, 109
- loi des grands nombres, 63

- macro-opérateurs, 201, 206
- MAE (mean average error), 691
- marge, 580
- marge d'un exemple, 587
- marge douce, 448
- marge géométrique, 443
- Markov
 - condition de, 359
 - couverture de, 364
 - équivalence de, 364
 - processus décisionnels de, 538
- Markov blanket*, 95–96
- matrice de codage des classes, 577
- matrice de confusion, 115
- matrice de Gram, 441
- matrice positive semi-définie, 435
- maximum a posteriori (MAP), 55
- maximum de vraisemblance, 50, 56
- Maximum Likelihood Estimator (MLE), 56
- MDLP (*Minimum Description Length Principle*), 52, 375, 497, 706–707, 759–761
- mélange de gaussiennes, 741
- membership query*, voir protocole de requêtes d'appartenance
- mémoire associative, 348
- mesure de dissimilarité, 597
- méta-apprentissage, 573
- méta-paramètre, 573
- méthodes à noyaux, 320, 429–486
- méthodes ε -gloutonnes, 550
- méthodes non paramétriques, 495, 509
- méthodes semi-paramétriques, 523
 - discrimination logistique, 523
 - mélange de distributions, 525
- méthodes spectrales, 610–611
- min-wise sampling*, 679
- minimisation du risque empirique, 49, 50
- minimisation du risque structurel (*SRM*), 712–713
- mixture models*, 319, 609–610
- modèle
 - sélection automatique, 702
 - modèle de mélange, 319, 609–610
 - modèle de Markov caché, voir HMM
 - modèles graphiques, 107
 - moindres carrés (estimation des), 302
 - Monte Carlo (méthode de), 547
 - mot engendré par une grammaire, 226
 - motif, 619
 - couverture de, 619
 - fréquence de, 619
 - support de, 619
 - motifs fréquents, 617–622
 - mots dérivés selon une grammaire, 226
 - multiplicateurs de Lagrange, 444, 709–710
 - mutation, 272, 273

 - négentropie, 627
 - neurone formel, 328
 - no-free-lunch theorem*, 74–78, 273
 - nombre de couverture, 701–702
 - non-negative matrix factorization (NMF)*, 630
 - non-paramétriques (méthodes), 495, 509
 - Novikoff
 - théorème de, 313
 - noyau
 - ANOVA, 463
 - arbres, 471
 - astuce des, 434
 - de convolution, 469
 - de Fisher, 477
 - diffusion, 474
 - gaussien, 464
 - génératif, 476–477
 - graphes, 471
 - reproduisant, 439
 - séquences, 468
 - spectre, 467
 - sur des ensembles, 465
 - sur des séquences, 466
 - sur des textes, 466
 - toutes sous-séquences, 469
 - universel, 459
 - noyaux (méthodes à), 320, 429–486

 - observation, 14
 - Occam (algorithme d'), 762–763
 - Occam (rasoir d'), 706
 - off-policy*
 - apprentissage par renforcement, 551
 - on policy*
 - apprentissage par renforcement, 550
 - one-class SVM*, 455
 - ontologie, 106
 - opérateurs de généralisation, 147–149
 - opérateurs de spécialisation, 148

- optimisation de performance, 17
- oracle, 14
- parcimonieux (codage), 632
- Pearson (corrélation de), 687
- performance (mesure de), 268
- pertinence du *MRE*, 66
- physique statistique (analyse de la), 720–723
- plan d'expériences, 668
- pointwise estimation*, 55
- politique, 531, 535
- politique optimale, 542
- POMDP (Partially Observable MDP), 561
- pool-based active learning*, 662
- poursuite de projections, 629
- prédiction, 15
- prédiction de séquence, 281
- prétraitement des données, 88
- présentation
 - complète, 234
 - négative, 234
 - positive, 234
- précision, 120
- pression sélective, 274
- principe de longueur minimale de description, 52, 375, 497, 706–707, 759–761
- principe de minimisation du risque structurel, 758
- problèmes trompeurs, 280
- processus décisionnels de Markov, 538
- processus de Markov, 386
- PRODIGY, 207
- PROGOL, 189–190
- programmation évolutive, 267
- programmation génétique, 267, 281–290
- programmation logique inductive (PLI), 161–197
- proie-prédateur, 292
- projection pursuit*, 629
- protocole d'apprentissage, 15
- protocole de requêtes d'appartenance, 663
- protocole de requêtes binaires arbitraires, 663
- protocole de requêtes d'équivalence, 663
- protocole de requêtes statistiques, 663
- pseudo-dimension, 702
- puissance du test, 116
- Q-learning*, 551
- query-by-committee*, 668
- raisonnement à partir de cas (RàPC), 211–212
- random forests*, 591
- ranking*, 60, 98
- rappel, 120
- recall*, 120
- reconnaissance des formes, xv
- redescription, 91
 - espace de, 433
- reformulation, 209
- règle de production d'une grammaire, 226
- règles d'association, 110–111, 617, 622–625, 679–680
 - confiance, 680
 - support, 679
- règles de contrôle, 201
- régression, 58–59, 301–302
 - aux moindres carrés, 495
 - de but, 203
- régression linéaire, 301–306
- régularisation
 - exemple, 420
 - théorie de la, 708–710
- remplacement, 276
- renforcement (signal de), 534, 535
- représentation duale, 432
- représentation par attribut-valeur, 109
- representer theorem*, 440
- réseaux bayésiens, 95, 107, 354–380
 - équivalence de Markov, 364
 - apprentissage des, 369–377
 - condition de Markov, 359
 - couverture de Markov, 364
 - d-séparation, 361–363, 368
 - diagnostic, 367
 - explication, 367
 - fonction de score, 374
 - méthode des arbres de jonction, 368
 - méthode EM, 372
 - méthodes de conditionnement, 368
 - méthodes de Monte Carlo, 368
 - raisonnement dans les, 365–368
- réseaux connexionnistes, 325–352
 - apprentissage, 334–345
 - architecture multicouche, 330
 - choix de l'architecture, 345
 - exemple de fonctionnement, 332
- réseaux de neurones
 - profonds, 346–347
- reservoir sampling*, 678
- reservoir computing*, 348–351
- resubstitution (estimation de l'erreur), 114
- résumés (calcul sur les flux), 679
- rétropropagation du gradient de l'erreur, 340
- ridge regression*, 320, 321
- risque empirique, 50
- risque réel, 45
- RKHS, voir espace de Hilbert à noyau reproduisant
- robot simulé, 288–290
- ROC (courbe), 121

- sac de mots, 465
 SARSA, 550
 sciences de l'artificiel, 5
 sélection, 273–276
 sélection (étape de), 274
 sélection automatique de modèle, 702
 sélection d'attributs, 90
 sélection de modèle, 52
 sélection par tournoi, 275
 sélection proportionnelle à la performance, 274
 sélection proportionnelle au rang, 275
sensitivity, 119
 séries temporelles, 349
 Shannon Claude, 11
 simplicité, 124
single representation trick, 145, 167
 SLD-résolution, 174
 SLD-subsumption, 179
small-scale learning, 719
 SOAR, 206
soft margin, 448
 Solomonoff (induction selon), 758
 sortie désirée, 14
 spécialisé maximalement général, 146
sparse coding, 632
 Spearman (corrélation de), 690
specificity, 119
speed-up learning, 205
 spline (régression), 318
S-set, 150
stacked generalization, 576–577
stepwise selection, 320
 stratégies d'évolution, 267, 281
stream-based active learning, 662
subgroup discovery, 625
 subsumption (relation de), 168
 support, voir règles d'association
 support (règle d'association), 623
 support d'un motif, 619
 sur politique
 apprentissage par renforcement, 550
 sur-complète (représentation), 632
 suradaptation, 9
 surapprentissage, 9
surrogate functions, 448
 SVM
 exemples critiques, 445
 lagrangien, 444
 mono-classe, 455
 multiplicateurs de Lagrange, 444
 sorties structurées, 456–458
 théorème de Mercer, 436
 variables duales, 444
 SVM transductifs (S3VMs), 647–649
 symétrisation (astuce de la), 700
 système apprenant, 14
 systèmes de classeurs, 293
 systèmes de recommandation, 686
 systèmes dynamiques, 723

 TD(λ), 551–552
 test
 de Friedman, 128
 de Wilcoxon, 127
 test multiple, 65
 test statistique
 score Z , 304
 tf-idf, 466, 690
 théorème central limite, 64
 théorème de Mercer, 436
 théorème de représentativité, 440
 théorème des schémas, 276–280
 trace d'éligibilité, 551
 transition de phase, 193–195
 treillis, 146
 tri, 60
 Turing Alan, 11

 UCB, 558
 UCT, 558–561
 utilité
 fonction d', 538, 541
 problème de l', 207

 validation croisée, 116–118
 Vapnik-Chervonenkis (dimension de), 699, 749, 751
 variables duales, 444
 variables latentes, 90, 377
 variance, 67
 vecteurs de support, voir exemples critiques
vector space model, 465
 vote, 576–577
 vraisemblance, 51
 fonction de, 370

weight decay, 710
 Wiener Norbert, 11
Wilcoxon signed rank test, 127