

Apprentissage statistique

**G. Dreyfus, J.-M. Martinez, M. Samuelides
M. B. Gordon, F. Badran, S. Thiria**

Sous la direction de Gérard Dreyfus

© Groupe Eyrolles, 2002, 2004, 2008,

ISBN : 978-2-212-12229-9

EYROLLES



Sommaire

Avant-propos et guide de lecture	XI
Guide de lecture	XII
Détail des contributions	XIII
1 L'apprentissage statistique : pourquoi, comment ?	1
Introduction	1
Premier exemple : un problème élémentaire d'apprentissage statistique	2
Point de vue algorithmique	3
Point de vue statistique	4
Quelques définitions concernant les modèles	5
Modèles statiques	5
Modèles dynamiques	6
Deux exemples académiques d'apprentissage supervisé	7
Un exemple de modélisation pour la prédiction	7
Un exemple de classification	11
Conclusion	16
Éléments de théorie de l'apprentissage	16
Fonction de perte, erreur de prédiction théorique	17
Dilemme biais-variance	22
De la théorie à la pratique	25
Remplacer des intégrales par des sommes	26
Bornes sur l'erreur de généralisation	27
Minimisation du risque structurel	30
Conception de modèles en pratique	30
Collecte et prétraitement des données	30
Les données sont préexistantes	30
Les données peuvent être spécifiées par le concepteur	30
Prétraitement des données	31
Sélection des variables	31
Apprentissage des modèles	32
Sélection de modèles	32
Sélection de modèles	32
Validation simple (hold-out)	32

Validation croisée (« cross-validation »)	33
Leave-one-out	34
Sélection de variables	35
Cadre théorique	36
Méthode de la variable sonde	37
Résumé : stratégies de conception	47
Conception de modèles linéaires par rapport à leurs paramètres (régression linéaire)	48
Sélection de variables pour les modèles linéaires en leurs paramètres	48
Apprentissage de modèles linéaires en leurs paramètres : la méthode des moindres carrés	49
Propriétés de la solution des moindres carrés	51
Estimation de la qualité de l'apprentissage	52
Interprétation géométrique	53
Dilemme biais-variance pour les modèles linéaires	54
Sélection de modèles linéaires	56
Moindres carrés par orthogonalisation de Gram-Schmidt	59
Éléments de statistiques	60
Qu'est-ce qu'une variable aléatoire ?	60
Espérance mathématique d'une variable aléatoire	62
Estimateur non biaisé	63
Variance d'une variable aléatoire	64
Autres distributions utiles	65
Intervalles de confiance	66
Tests d'hypothèse	68
Conclusion	70
Bibliographie	70

2 Les réseaux de neurones 73

Introduction	73
Réseaux de neurones : définitions et propriétés	73
Les neurones	74
Les réseaux de neurones	75
Propriété fondamentale des réseaux de neurones statiques (non bouclés) : l'approximation parcimonieuse	82
À quoi servent les réseaux de neurones non bouclés à apprentissage supervisé ?	
Modélisation statique et discrimination (classification)	84
À quoi servent les réseaux de neurones à apprentissage non supervisé ? Analyse et visualisation de données	87

À quoi servent les réseaux de neurones bouclés à apprentissage supervisé ? Modélisation dynamique « boîte noire » et « semi-physique » ; commande de processus	87
Quand et comment mettre en œuvre des réseaux de neurones à apprentissage supervisé ?	88
Quand utiliser les réseaux de neurones ?	88
Comment mettre en œuvre les réseaux de neurones ?	89
Conclusion	93
Réseaux de neurones à apprentissage supervisé et discrimination (classification)	93
Quand est-il opportun d'utiliser un classifieur statistique ?	93
Classification statistique et formule de Bayes	95
Classification et régression	96
Modélisation et classification de données structurées : les « graph machines »	103
Définitions	104
Apprentissage	105
Deux exemples académiques	106
Exemples d'applications	107
Introduction	107
Reconnaissance de formes :	
la lecture automatique de codes postaux	107
Une application en contrôle non destructif :	
la détection de défauts dans des rails par courants de Foucault	111
Fouille de données : le filtrage de documents	112
Aide à la découverte de médicaments : prédiction de propriétés chimiques et d'activités thérapeutiques de molécules	116
Une application en formulation :	
la prédiction de la température de liquidus de verres	118
Modélisation d'un procédé de fabrication : le soudage par points	118
Application en robotique :	
modélisation de l'actionneur hydraulique d'un bras de robot	121
Modélisation semi-physique d'un procédé manufacturier	122
Contrôle de l'environnement : hydrologie urbaine	123
Une application en robotique mobile :	
le pilotage automatique d'un véhicule autonome	124
Techniques et méthodologie de conception de modèles statiques (réseaux non bouclés)	125
Sélection des variables	126
Estimation des paramètres (apprentissage) d'un réseau de neurones non bouclé	126

Sélection de modèles	143
Techniques et méthodologie de conception de modèles dynamiques (réseaux bouclés ou récurrents)	156
Représentations d'état et représentations entrée-sortie	157
Les hypothèses concernant le bruit et leurs conséquences sur la structure, l'apprentissage et l'utilisation du modèle	158
Apprentissage non adaptatif des modèles dynamiques sous forme canonique	166
Que faire en pratique ? Un exemple réel de modélisation « boîte noire »	172
Mise sous forme canonique des modèles dynamiques	175
Modélisation dynamique « boîte grise »	179
Principe de la modélisation semi-physique	179
Conclusion : quels outils ?	188
Compléments théoriques et algorithmiques	189
Quelques types de neurones usuels	189
Algorithme de Ho et Kashyap	191
Complément algorithmique : méthodes d'optimisation de Levenberg-Marquardt et de BFGS	191
Complément algorithmique : méthodes de recherche unidimensionnelle pour le paramètre d'apprentissage	193
Complément théorique : distance de Kullback-Leibler entre deux distributions gaussiennes	194
Complément algorithmique : calcul des leviers	196
Bibliographie	197

3 Compléments de méthodologie pour la modélisation : réduction de dimension et ré-échantillonnage

Pré-traitements	204
Pré-traitements des entrées	204
Pré-traitement des sorties pour la classification supervisée	204
Pré-traitement des sorties pour la régression	205
Réduction du nombre de composantes	206
Analyse en composantes principales	206
Principe de l'ACP	206
Analyse en composantes curvilignes	210
Formalisation de l'analyse en composantes curvilignes	211
Algorithme d'analyse en composantes curvilignes	212
Mise en œuvre de l'analyse en composantes curvilignes	213
Qualité de la projection	214
Difficultés présentées par l'analyse en composantes curvilignes	214

Application en spectrométrie	215
Le bootstrap et les réseaux de neurones	216
Principe du bootstrap	217
Algorithme du bootstrap pour calculer un écart-type	218
L'erreur de généralisation estimée par bootstrap	218
La méthode NeMo	219
Test de la méthode NeMo	221
Conclusions	223
Bibliographie	224
4 Identification « neuronale » de systèmes dynamiques commandés et réseaux bouclés (récurrents)	225
<hr/>	
Formalisation et exemples de systèmes dynamiques commandés à temps discret	226
Formalisation d'un système dynamique commandé par l'équation d'état	226
Exemple d'un système dynamique à espace d'état discret	227
Exemple d'un oscillateur linéaire	227
Exemple du pendule inversé	228
Exemple d'un oscillateur non linéaire : l'oscillateur de Van der Pol	229
Introduction d'un bruit d'état dans un système dynamique à espace d'état discret : notion de chaîne de Markov	229
Introduction d'un bruit d'état dans un système dynamique à états continus : modèle linéaire gaussien	231
Modèles auto-régressifs	231
Limites des modélisations des incertitudes sur le modèle par un bruit d'état	233
Identification de systèmes dynamiques commandés par régression	233
Identification d'un système dynamique commandé par régression linéaire	233
Identification d'un système dynamique non linéaire par réseaux de neurones non bouclés	237
Identification adaptative (en ligne) et méthode de l'erreur de prédiction réursive	239
Estimateur récursif de la moyenne empirique	239
Estimateur récursif de la régression linéaire	241
Identification réursive d'un modèle AR	242
Méthode générale de l'erreur de prédiction réursive	243
Application à l'identification neuronale d'un système dynamique commandé	244
Filtrage par innovation dans un modèle d'état	245
Introduction d'une équation de mesure et problème du filtrage	245
Filtrage de Kalman	247
Extension du filtre de Kalman	251

Apprentissage adaptatif d'un réseau de neurones par la méthode du filtrage de Kalman	252
Réseaux neuronaux récurrents ou bouclés	254
Simulateur neuronal d'un système dynamique commandé en boucle ouverte	254
Simulateur neuronal d'un système dynamique commandé en boucle fermée	255
Quelques réseaux bouclés particuliers	255
Mise sous forme canonique des réseaux bouclés	258
Apprentissage des réseaux de neurones récurrents ou bouclés	258
Apprentissage dirigé (teacher forcing)	259
Dépliage de la forme canonique et rétropropagation à travers le temps	260
Apprentissage en temps réel des réseaux bouclés	262
Application des réseaux neuronaux bouclés à l'identification de systèmes dynamiques commandés mesurés	263
Compléments algorithmiques et théoriques	264
Calcul du gain de Kalman et propagation de la covariance	264
Importance de la distribution des retards dans un réseau récurrent	266
Bibliographie	267
5 Apprentissage d'une commande en boucle fermée	269
<hr/>	
Généralités sur la commande en boucle fermée des systèmes non linéaires	269
Principe de la commande en boucle fermée	269
Commandabilité	270
Stabilité des systèmes dynamiques commandés	271
Synthèse d'une commande « neuronale » par inversion du modèle du processus	273
Inversion directe	273
Utilisation d'un modèle de référence	276
Commande avec modèle interne	277
Commande prédictive et utilisation des réseaux récurrents	278
Programmation dynamique et commande optimale	280
Exemple de problème déterministe à espace d'états discret	280
Exemple de problème de décision markovienne	281
Définition d'un problème de décision markovienne	282
Programmation dynamique à horizon fini	286
Programmation dynamique à horizon infini et à coût actualisé	287
Problèmes de décision markovienne partiellement observés	288
Apprentissage par renforcement et programmation neuro-dynamique	289
Évaluation d'une politique par la méthode de Monte-Carlo et apprentissage par renforcement	289

Présentation de l'algorithme TD d'évaluation d'une politique	290
Apprentissage par renforcement : méthode du Q-learning	292
Apprentissage par renforcement et approximation neuronale	294
Bibliographie	297
6 La discrimination	301
Apprentissage de la discrimination	302
Erreurs d'apprentissage et de généralisation	303
Surfaces discriminantes	304
Séparation linéaire : le perceptron	305
Géométrie de la classification	306
Algorithmes d'apprentissage pour le perceptron	309
Algorithme Minimerror	317
Exemple d'application : la classification de signaux de sonar	318
Algorithmes d'apprentissage adaptatifs (« en ligne »)	320
Interprétation de l'apprentissage en termes de forces	320
Au-delà de la séparation linéaire	321
Perceptron sphérique	321
Heuristiques constructives	322
Algorithme constructif NetLS	323
Machines à vecteurs supports (Support Vector Machines)	325
SVM à marge dure	327
Machines à noyaux (Kernel machines)	329
SVM à marge floue (Soft margin SVM)	331
SVM pratique	333
Problèmes à plusieurs classes	334
Questions théoriques	335
Formulation probabiliste de l'apprentissage et inférence bayésienne	335
Théorie statistique de l'apprentissage	340
Prédiction du comportement typique des classifieurs	342
Compléments	344
Bornes du nombre d'itérations de l'algorithme du perceptron	344
Nombre de dichotomies linéairement séparables	345
Bibliographie non commentée	345
7 Cartes auto-organisatrices et classification automatique	349
Notations et définitions	351

Méthode des k-moyennes	352
Présentation de l'algorithme	352
Version stochastique des k-moyennes	354
Interprétation probabiliste des k-moyennes	357
Carte topologique auto-organisatrice	360
Les cartes auto-organisatrices	360
L'algorithme d'optimisation non adaptative des cartes topologiques	363
L'algorithme de Kohonen	369
Discussion	370
Architecture neuronale et carte topologique	371
Architecture et carte topologique évolutive	372
Interprétation de l'ordre topologique	373
Carte topologique probabiliste	375
Classification et carte topologique	378
Étiquetage de la carte par données expertisées	378
Recherche d'une partition adaptée aux classes recherchées	379
Étiquetage et classification	381
Applications	382
Une application en télédétection satellitaire	383
Carte topologique et recherche documentaire	407
Extension des cartes topologiques aux données catégorielles	409
Codage et analyse des données catégorielles	409
Cartes topologiques et données binaires	410
Cartes topologiques probabilistes et données catégorielles (CTM)	413
Discussion	416
Exemples d'application	417
Le modèle BTM	417
Analyse des correspondances multiples	418
Le modèle CTM	419
Bibliographie	424

Bibliographie commentée	427
--------------------------------	-----

Outils pour les réseaux de neurones et contenu du CD-Rom	431
Installer Neuro One	431
Présentation des exemples	436
Exemple 1	436

Exemple 2	436
Exemple 3	437
Exemple 4	437
Exemple 5	437
Installation des exemples	437
Compiler le code source	438
Exécuter le code source	438
Exécuter le code source Visual Basic	439
Visualiser les modèles	440
La librairie NDK (Neuro Developer Kit)	440
Programme de démonstration de la librairie	440
Les compilateurs C	441
Licence	442
Index	443
