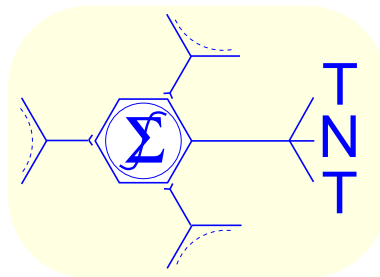

Réseaux de neurones

Théories et applications :
automates cellulaires,
biologie,

Hopfield :
mémoires associatives
et optimisation

Daniel Collobert
FT-R&D Lannion



Sommaire

-1	Avant-propos	1
-1.1	Liminaire	1
-1.1.1	Du point de vue de l'industrie	2
-1.1.2	Du point de vue académique	3
-1.1.3	Quatre applications	3
0	Mythes	5
0.1	Minsky	5
1	Biologie	7
1.1	Le cerveau	8
1.1.1	Historique	8
1.1.2	Méthodes d'étude	8
1.1.3	Structure	8
1.1.4	Fonction	8
1.2	Le neurone	9
1.2.1	Historique	10
1.2.2	Méthodes d'étude	10
1.2.3	Fonction	13
1.2.4	Structure	13
1.2.5	L'influx nerveux	13
1.3	Les synapses	17
1.3.1	Historique	17
1.3.2	Structure	17
1.3.3	Les neurorégulateurs	18
1.3.4	La névroglie	20
1.4	Génèse	22
1.5	Perception et Mémoire	23
1.6	L'esprit	24
1.7	la conscience	25

2 Mémoires associatives	27
2.1 Historique	28
2.2 Introduction	28
2.3 Première approche : une mémoire hétéro-associative	29
2.3.1 Préliminaires : le neurone formel	29
2.3.2 Préliminaires : l'apprentissage de Hebb	30
2.3.3 Objectifs et méthode	31
2.3.4 Exercice	31
2.4 Seconde approche : une mémoire auto-associative récurrente	34
2.4.1 Préliminaires	34
2.4.2 Formalisation	35
2.4.3 Mémoriser un motif	35
2.4.4 Mémoriser plusieurs motifs	36
2.4.5 Capacité de mémorisation	36
2.4.6 Capacité de stockage du réseau de Hopfield	37
2.4.7 Fonction énergie	39
2.4.8 États spurieux	42
2.4.9 Réseau de Hopfield et physique des verres de spins	43
2.4.10 Prototypes corrélés : la pseudo-inverse	45
2.5 Résultats de simulation	46
2.6 Implémentations optiques	46
2.7 Troisième approche : réseaux de Hopfield et Optimisation	48
2.7.1 Modifications du modèle original	48
2.7.2 Modèle continu	48
2.7.3 Énergie	49
2.7.4 Renormalisation	50
2.7.5 Application au convertisseur analogique–numérique	51
2.7.6 Application au problème du voyageur de commerce	53
2.7.7 Application au problème des bibliothécaires	56
2.7.8 Application à l'imagerie médicale	56
2.8 Conclusions	58
2.9 Exercices	58
2.10 Compléments : systèmes dynamiques	58
3 Apprentissage	59
3.1 Processus d'apprentissage	61
3.1.1 Apprentissage par correction d'erreur	61
3.1.2 Apprentissage de Hebb	62
3.1.3 Apprentissage compétitif	63
3.1.4 Apprentissage de Boltzmann	63
3.1.5 Le "crédit assignment problem"	64

3.1.6	Apprentissage supervisé	64
3.1.7	Apprentissage par renforcement	65
3.1.8	Apprentissage non supervisé	65
3.1.9	Différentes tâches d'apprentissage	65
3.1.10	Adaptation et apprentissage	66
3.2	Apprentissage et statistique	67
3.2.1	L'apprentissage dans les RN est un processus de régression linéaire	67
3.2.2	Dilemme biais/variance pour la généralisation	68
3.3	Théorie de l'apprentissage	71
3.3.1	Position du problème	71
3.3.2	Minimisation du risque empirique	72
3.3.3	VCdim	74
3.3.4	Taux de convergence uniforme	74
3.3.5	A faire, peut être à placer après	76
3.3.6	Minimisation de l'erreur structurelle	76
3.3.7	Contrôle de la capacité	77
3.4	Discussion	77
3.5	Retour à la biologie	77
4	KOHONEN	79
4.1	Introduction	79
4.2	Préliminaires	79
4.2.1	Tessellation de Voronoï	80
4.2.2	Triangulation de Delaunay	81
4.2.3	Utilisation de la tessellation de Voronoï	81
4.3	Algorithme de Kohonen	81
4.3.1	Minimiser une erreur au sens de Bayes	81
4.3.2	Minimiser une erreur sur un critère de distance	82
4.3.3	Utiliser un espace de représentation de petite dimension	82
4.3.4	Cartes auto-organisatrices de Kohonen	83
4.3.5	Exemple	83
4.3.6	Remarques	84
4.4	Learning Vector Quantization (LVQ)	84
4.4.1	LVQ1	85
4.4.2	LVQ2	85
4.4.3	LVQ3	86
4.5	Conclusions	86
4.6	Notes	86

6 Réseaux multicouches	87
6.1 Systèmes paramétriques et non paramétriques	87
6.2 Un rapide survol avant de commencer	88
6.3 Notion de neurone formel	89
6.4 Propriétés d'un neurone	89
6.5 Le Perceptron	90
6.5.1 Le talon d'Achille du Perceptron	91
6.6 Apprentissage des poids d'un neurone formel	92
6.6.1 Limitations	93
6.6.2 Perspectives	93
6.7 Minimisation d'une fonction de coût	93
6.7.1 exemples de fonction de coût	94
6.7.2 procédure de minimisation	94
6.8 Bref historique	95
6.9 Réseaux de neurones : complexité	95
6.10 NOTES	95
6.11 Régression linéaire	96
6.12 Unités non linéaires	97
6.13 Méthode historique : première version	98
6.14 Méthode historique : deuxième version	99
6.15 Méthode du Lagrangien	99
6.16 Amélioration : autre fonction de coût	99
6.17 Amélioration : ajout d'un terme d'inertie	99
6.18 Méthode de Gauss-Newton	99
6.19 Méthode de Levenberg-Marquard	99
6.20 Amélioration : pentes optimisées	99
6.21 Relations entre pente et pas	102
6.22 Régularisation : weight elimination	103
6.23 Régularisation : weight decay	106
6.24 A propos des données	107
6.24.1 Bagging	108
6.24.2 Boosting	109
6.24.3 Biais et variance	110
6.25 Apprentissage statistique et généralisation dans les MLP	111
6.25.1 Minimisation de l'erreur et maximum de vraisemblance	111
6.25.2 Distribution de Gibbs	112
7 TP	115
7.1 Introduction	115
7.2 Rappels	117
7.3 Application au XOR	118
7.4 Un encodeur 4-2-4	119

9	À propos de la couleur	123
9.1	Première expérience	124
9.2	Deuxième expérience	124
9.3	La couleur du point de vue physique	124
9.3.1	La lumière	124
9.3.2	Les sources de lumière	124
9.3.3	La température de couleur	125
9.3.4	Efficacité lumineuse spectrale	126
9.3.5	Les unités	127
9.4	La vision des couleurs	128
9.4.1	Le récepteur visuel humain	129
9.4.2	Vision diurne et vision nocturne	130
9.4.3	Théories de la vision des couleurs	130
9.4.4	Anomalies de la vision colorée	132
9.5	Colorimétrie	133
9.5.1	Couleurs du spectre	133
9.5.2	Trivariance visuelle	134
9.5.3	Lumières simples, complexes, complémentaires	136
9.6	La couleur selon Munsell	137
9.6.1	Un langage universel des couleurs	138
9.7	La trichromie	140
9.7.1	Système trichromique RVB (ou RGB)	141
9.7.2	Calcul des composantes trichromatiques	144
9.8	Quelques couleurs	144
9.9	Conclusions	146
8	Les mélanges d'experts et l'algorithme EM	147
8.1	Introduction	147
8.2	La théorie du modèle modulaire	147
8.3	L'architecture du modèle modulaire	150
8.4	L'algorithme EM (Expectation-Maximization)	151
8.5	Application	155
8.6	Comparaison avec une autre fonction de coût	158
8.7	Convergence de l'algorithme EM	160
8.8	Jensen et Fisher	162
8.8.1	Inégalité de Jensen	162
8.8.2	Formule de Fisher	163
9	Introduction	165

10 Théorie de l'apprentissage statistique	167
10.1 Apprendre avec un nombre limité d'exemples	167
10.2 Un problème de convergence uniforme	168
10.3 Le dilemme biais-variance	170
10.4 Mesure de la qualité d'une solution	172
11 Support Vector Machines	173
11.1 Un problème de classification	173
11.2 Une première idée de solution	173
11.3 À la recherche de l'hyperplan optimal dans le cas séparable	175
11.4 Une autre idée de solution	177
11.5 Une idée géniale	179
11.6 Construction de <i>kernels</i>	181
11.7 Problèmes existentiels	182
11.8 Un petit exemple : le XOR	184
12 Support Vector Regression	189
13 SVM<i>Torch</i>	193
14 Support Vector Machines for Large-Scale Regression Problems	195
14.1 Introduction	195
14.2 The Decomposition Algorithm	196
14.2.1 Selection of a New Working Set	197
14.2.2 Solving the Sub-Problem	198
14.2.3 Shrinking	201
14.2.4 Termination Criterion	202
14.2.5 Implementation Details	202
14.3 Experimental Results	203
14.3.1 Speed Comparisons	203
14.3.2 Analysis of the Cache Size	203
14.4 Conclusion	205
14.5 <i>Appendix</i> : Things that should have been in the paper	206
14.5.1 On the experiments	206
14.5.2 On the decomposition method	206
14.5.3 On the convergence of the algorithm	206
14.6 Remarks on the relation between many SVM algorithms	207
14.7 Conclusion	208
15 On the Convergence of SVM<i>Torch</i>	211
15.1 Introduction	211
15.2 The Convergence Theorem of Keerthi and Gilbert	214
15.3 Convergence of our Algorithm	215
15.4 Conclusion	216

16 Expériences	217
16.1 Présentation du problème	217
16.2 La capacité en pratique	218
16.3 État de l'Art ?	220
17 Épilogue	223
Vitamines	223
A Théorie de l'information	225
A.1 Préliminaires	225
A.1.1 Espace de probabilité	225
A.1.2 Application mesurable	226
A.1.3 Variable aléatoire	226
A.1.4 Quelques autres définitions et propriétés des probabilités	226
A.1.5 Signification de la probabilité	227
A.1.6 Information associée à un événement	228
A.1.7 Information mutuelle associée à deux événements	228
A.2 Entropie et autres grandeurs liées à la quantité d'information	228
A.2.1 Autres définitions utiles	231
A.3 Exercices	232
A.4 Corrigé des exercices	233
B Complexité	239
B.1 Notions de complexité	239
B.2 La classe P	240
B.3 La classe NP	241
B.4 La classe NP -complet	241
C Petite introduction à la programmation non-linéaire	243

