

THESE

présentée à

L'UNIVERSITE PIERRE ET MARIE CURIE (PARIS VI)

pour obtenir le titre de

DOCTEUR DE L'UNIVERSITE PARIS VI

Spécialité :

MATHEMATIQUES

par

Abdallah MKHADRI

Sujet de la thèse :

**CLASSIFICATION ET DISCRIMINATION DES
DONNEES QUALITATIVES
DISCRIMINATION MULTINOMIALE REGULARISEE**

Soutenue le 21 Décembre 1990 devant le Jury composé de :

MM.	Paul	DEHEUELS	Président
	Gérard	GOVAERT	Rapporteur
	François	MARCOTORCHINO	Rapporteur
	Michel	BRONIKOWSKI	
	Gilles	CELEUX	
	Edwin	DIDAY	Examineurs
	Gilbert	SAPORTA	

TABLE DES MATIERES

CHAPITRE 1

PONDERATION DES VARIABLES POUR LA CLASSIFICATION BINAIRE AVEC CRITERE DE TYPE INERTIE INTRA-CLASSES

INTRODUCTION	11
1. INERTIE SUR DONNÉES BINAIRES	12
2. ALGORITHMES NON ADAPTATIFS	15
2.1 L'algorithme MNDBIN	16
2.2 L'algorithme MNDBVP	17
3. ALGORITHMES ADAPTATIFS	20
3.1 Contrainte multiplicative	22
3.2 Contrainte additive	23
3.3 Contrainte logistique	24
4. APPROCHE PROBABILISTE	25
5. APPLICATIONS ET COMPARAISON DES MÉTHODES	28
5.1 Données simulées	29
5.2 Données DKS	29
6. PROBLEMES DE DÉGÉNÉRESCENCE	33
6.1 Contrainte multiplicative	33
6.2 Contrainte additive	34

6.3	Contrainte logistique	35
6.4	Cas "modèle"	35
7.	DISCUSSION	36
	ANNEXE 1	41

CHAPITRE 2		
DISCRIMINATION BINAIRE		
	INTRODUCTION	47
1.	NOTATIONS ET REGLE DE DÉCISION	48
2.	MODELES PARAMÉTRIQUE	50
2.1	Modèles basés sur l'estimation des densités conditionnelles	50
2.1.1	Modèle multinomial complet	50
2.1.2	Modèle d'indépendance du premier ordre	51
2.1.3	Procédures des plus proches voisins	52
2.2	Modèle logistique linéaire	53
2.2.1	Introduction	53
2.2.2	Méthodes d'échantillonnage et fonction de vraisemblance	54
2.2.3	Estimateurs du maximum de vraisemblance	55
2.2.4	Méthode de Newton-Raphson	56
3.	LE MODELE NON PARAMÉTRIQUE	56
3.1	Introduction	58

3.2 Méthodes d'estimation du paramètre de lissage	60
4. AUTRES PROCÉDURES	60
4.1 Procédures utilisant les polynômes orthogonaux	60
4.1.1 Modèle de Martin - Bradley	61
4.1.2 Modèle de Kronmal-Ott-Tarter	62
4.2 Distance de Matusita	65
4.3 Le modèle loglinéaire	67
4.5 Discrimination par arbre binaire	67
5. SIMULATIONS ET COMPARAISON DES MÉTHODES	68
5.1 Données digitales	70
5.2 données store	74
CONCLUSION	77

CHAPITRE 3

TRAITEMENT COMPLET D'UN PROBLEME DE DISCRIMINATION STATISTIQUEMENT MAL POSE

1. INTRODUCTION	79
2. PRESENTATION DES DONNEES	79
3. LE PROBLEME STATISTIQUE	80
4. TECHNIQUES MULTIDIMENSIONNELLES DE RÉDUCTION DE LA DIMENSION	81
4.1 Codage des variables qualitatives	81

4.2 Discrimination barycentrique	81
5. SÉLECTION DE VARIABLES ET CLASSIFICATION	82
6. DISCRIMINATION BINAIRE PAR LA MÉTHODE DES NOYAUX	84
7. DISCUSSION DES RÉSULTATS	86
8. CONCLUSION	88

CHAPITRE 4

DISCRIMINATION BINAIRE NON PARAMETRIQUE METHODES D'ESTIMATION DU PARAMETRE DE LISSAGE

INTRODUCTION	91
1. PROCÉDURES DE LISSAGE DANS LE CADRE DE L'ESTIMATION DE LA DENSITÉ	93
1.1 Propriétés de la méthode du maximum de vraisemblance	93
1.2 Méthode de Hall et critique de la validation croisée	94
1.3 Autres procédures	98
1.3.0 Pseudo-vraisemblance	98
1.3.1 Une procédure de validation croisée	100
1.3.2 Une procédure de vraisemblance classifiante	103
1.3.3 Procédure basée sur le Bootstrap	104
2 MÉTHODES D'ESTIMATION LIÉES À LA DISCRIMINATION	109
2.1 Méthode de Tutz	110
2.2 Méthode de Hall & Wand	114
3. APPLICATION PRATIQUE	118

4. DONNÉES QUALITATIVES ET TABLES DE CONTINGENCE	122
5. CONCLUSION	125
ANNEXE 4	127

CHAPITRE 5

DISCRIMINATION MULTINOMIALE REGULARISEE

1. INTRODUCTION	133
2. DISCRIMINATION GAUSSIENNE RÉGULARISER	135
3. LA DISCRIMINATION MULTINOMIALE REGULARISEE	137
3.1 Schéma de régularisation	138
3.2 Le paramètre de complexité	139
3.2 Les paramètres optimaux de lissage	145
4. APPLICATION SUR DONNÉES RÉELLES	149
5. SELECTION DE VARIABLES ET RÉGULARISATION	155
6. CONCLUSION ET REMARQUES	156
ANNEXE 5	158

BIBLIOGRAPHIE	161
----------------------	-----