

## Table des matières

Introduction .....	1
Éléments d'histoire de l'Analyse factorielle des données.....	4
Contenu du livre .....	6
Analyse en composantes principales.....	7
1.1. Introduction .....	7
1.2. Données .....	8
1.3. Analyse du nuage des individus.....	10
1.3.1. Ajustement du nuage des individus par une droite.....	10
1.3.2. Ajustement du nuage des individus par un plan.....	13
1.3.3. Ajustement du nuage des individus par un sous-espace.....	16
1.4. Analyse du nuage des variables .....	17
1.5. Relation entre l'analyse des individus et l'analyse des variables.....	18
1.6. Reconstitution du tableau de données.....	20
1.7. Variantes de l'analyse en composantes principales.....	22
1.7.1. Analyse générale .....	22
1.7.2. Centrage du nuage des individus .....	22
1.7.3. ACP normée .....	24
1.8. Interprétation des résultats .....	27
1.8.1. Outils pour l'interprétation.....	27
1.8.1.1. Apport d'un axe factoriel .....	27
1.8.1.2. Choix du nombre d'axes .....	29
1.8.1.3. Influence d'un individu ou d'une variable.....	31
1.8.2. Variables et individus supplémentaires .....	31
1.8.3. Règles d'interprétation .....	34

1.9. Exemples .....	40
1.9.1. Analyse des données concernant les mammifères .....	40
1.9.2. Analyse des données « Textures » .....	41
1.10. Compléments .....	46
1.10.1. Distance générale et pondération différenciée des individus .....	46
1.10.2. Robustesse de l'analyse et codage par les rangs .....	49
1.10.3. Courte introduction aux réseaux de neurones formels .....	50
1.10.4. Relation entre ACP et perceptrons multicouches .....	52
1.10.5. Analyse non linéaire basée sur des fonctions noyau .....	55
1.11. Exercices .....	64
Exercice 1.1 .....	64
Exercice 1.2 .....	66
Exercice 1.3 .....	70
Exercice 1.4 .....	73
Exercice 1.5 .....	75
Exercice 1.6 .....	76
Analyse factorielle des correspondances .....	81
2.1. Introduction .....	81
2.2. Présentation de la méthode .....	82
2.2.1. Données .....	82
2.2.2. Pondération des profils .....	88
2.2.3. Évaluation de la distance entre des profils .....	89
2.2.4. Analyse du nuage des profils-lignes .....	92
2.2.5. Analyse du nuage des profils-colonnes .....	97
2.2.6. Relations entre les analyses des deux nuages .....	99
2.2.7. Présence de variables quantitatives .....	103
2.3. Interprétation des résultats .....	105
2.3.1. Outils pour l'interprétation .....	105
2.3.1.1. Test d'indépendance entre les variables .....	105
2.3.1.2. Apport d'un axe factoriel .....	107
2.3.1.3. Choix du nombre d'axes .....	108
2.3.1.4. Influence de chaque profil .....	108
2.3.2. Profils supplémentaires .....	110
2.3.3. Règles d'interprétation .....	112
2.4. Compléments .....	114
2.4.1. Analyse du nuage centré et analyse du nuage non centré .....	114
2.4.2. Analyse directe des profils .....	117
2.5. Exemples .....	120
2.5.1. Enquête « Internet : accès et utilisation au Québec » .....	120
2.5.1.1. Croisement âge – services .....	120

2.5.1.2. Croisement âge – ancienneté .....	124
2.5.2. Analyse d'un tableau de notes .....	126
2.6. Exercices.....	132
Exercice 2.1 .....	132
Exercice 2.2 .....	133
Exercice 2.3 .....	134
Analyse des correspondances multiples.....	135
3.1. Introduction .....	135
3.2. Présentation de la méthode.....	136
3.2.1. Données .....	136
3.2.2. Analyse du tableau disjonctif complet.....	140
3.2.2.1. Analyse du nuage des individus et du nuage des modalités.....	140
3.2.2.2. Relations entre le nuage des individus et le nuage des modalités....	148
3.2.2.3. Représentation synthétique des variables .....	149
3.2.3. Analyse du tableau de Burt .....	150
3.2.4. Présence de variables quantitatives.....	152
3.3. Interprétation des résultats .....	153
3.3.1. Outils pour l'interprétation.....	153
3.3.1.1. Apport d'un axe factoriel .....	153
3.3.1.2. Influence de chaque profil.....	155
3.3.1.3. Influence de chaque variable .....	156
3.3.2. Individus et variables supplémentaires .....	158
3.3.3. Règles d'interprétation .....	161
3.4. Exemples .....	165
3.4.1. Enquête « Les étudiants et la ville » .....	165
3.4.2. Enquête « Internet : accès et utilisation au Québec ».....	169
3.5. Exercices.....	175
Exercice 3.1 .....	175
Exercice 3.2 .....	176
Exercice 3.3 .....	176
Analyse factorielle discriminante.....	179
4.1. Introduction .....	179
4.2. Analyse factorielle discriminante linéaire .....	180
4.2.1. Notations et support théorique .....	180
4.2.1.1. Centre de gravité de l'échantillon et centres de gravité des classes.	181
4.2.1.2. Covariance entre deux variables descriptives.....	182
4.2.1.3. Covariance intra-classe et covariances interclasses .....	182
4.2.2. Étape descriptive .....	184
4.2.2.1. Calculs préliminaires.....	184
4.2.2.2. Recherche du premier axe factoriel discriminant.....	186

4.2.2.3. Recherche des autres axes factoriels discriminants.....	188
4.2.2.4. Traitement algorithmique alternatif.....	189
4.2.2.5. Cas particulier des matrices de covariances mal conditionnées.....	191
4.2.2.6. Pondération différenciée des observations.....	192
4.2.2.7. Cas particulier de deux groupes.....	193
4.2.2.8. Lien entre l'AFD et l'ACP.....	193
4.2.2.9. Lien entre l'AFD et la régression linéaire.....	194
4.2.3. Étape décisionnelle.....	195
4.2.3.1. Cas général.....	195
4.2.3.2. Cas particulier de deux groupes.....	196
4.2.4. Nombre d'axes factoriels discriminants significatifs.....	196
4.2.5. Exemple.....	200
4.3. Analyse factorielle discriminante quadratique.....	202
4.3.1. Support théorique.....	203
4.3.1.1. Métrique diagonale.....	204
4.3.1.2. Métrique quelconque de déterminant unitaire.....	207
4.4. Compléments.....	208
4.4.1. Relation entre AFD et perceptrons multicouches.....	208
4.4.2. Analyse discriminante à noyaux.....	214
4.4.3. Analyse discriminante asymétrique.....	221
4.5. Exercices.....	226
Exercice 4.1.....	226
Exercice 4.2.....	226
Exercice 4.3.....	227
Exercice 4.4.....	229
Exercice 4.5.....	232
Solutions des exercices.....	235
Chapitre 1.....	235
Exercice 1.1.....	235
Exercice 1.2.....	239
Exercice 1.3.....	242
Exercice 1.4.....	247
Exercice 1.5.....	249
Exercice 1.6.....	250
Chapitre 2.....	252
Exercice 2.1.....	252
Exercice 2.2.....	253
Exercice 2.3.....	253
Chapitre 3.....	255
Exercice 3.1.....	255

Exercice 3.2 .....	256
Exercice 3.3 .....	261
Chapitre 4 .....	264
Exercice 4.1 .....	264
Exercice 4.2 .....	265
Exercice 4.3 .....	267
Exercice 4.4 .....	268
Exercice 4.5 .....	269
Bibliographie .....	273
Table des matières .....	ii
Index .....	vii

## Index

- ACM, 135
- ACP, 7
  - centrée, 22
  - générale, 22
  - normée, 24
- ADA, 221
- AFC, 81
- AFCB, 81
- AFD, 179
  - décisionnelle, 195
  - descriptive, 184
- Aides à l'interprétation, 27, 153
- Aides à l'interprétation, 105
- Ajustement
  - des individus, 10
  - des variables, 17
  - par un plan, 13
  - par un sous-espace, 16
  - par une droite, 10
- Analyse
  - canonique, 177
  - des correspondances multiples, 135
  - des modalités, 140
  - des profils-colonnes, 97
  - des profils-lignes, 92
  - du tableau de Burt, 150
  - du tableau disjonctif complet, 140
  - en composantes principales, 7
    - à noyaux, 56
  - exploratoire, 7
  - factorielle des correspondances, 81
  - factorielle discriminante, 179
    - à noyaux, 214
    - asymétrique, 221
    - linéaire, 180
    - quadratique, 202
  - individus, 10
  - profils initiaux, 117
  - tableau de notes, 126
  - variables, 17
- Apprentissage du PMC, 52
- Auto-association, 53
- Axe factoriel
  - deuxième, 15
  - discriminant, 186
  - premier, 13
- Barycentre, 22
- Base orthonormée, 16

- Centre de gravité, 22, 88, 143, 181
- Codage condensé, 136
- Coefficient de corrélation, 25
- Compression de données, 21
- Conditionnement, 191
- Connexion, 51
- Contribution
  - absolue, 31, 108, 155
  - relative, 27, 107, 154
- Coordonnée
  - d'un profil-colonne, 98
  - d'un profil-ligne, 95
  - de point-individu, 13
  - de point-variable, 18
- Cosinus carré, 28, 107
- Couche
  - cachée, 51
  - d'entrée, 51
  - de sortie, 51
- Covariance
  - empirique, 182
  - interclasses, 183
  - intra-classes, 183
- Dérivée partielle, 12
- Distance
  - de Mahalanobis, 195
  - du Chi-2, 90, 143
  - euclidienne usuelle, 11, 17
  - générale, 46
- Ecart-type, 24
- Echantillon test, 199
- Eckart-Young, 21
- Effet taille, 45
- Equivalence distributionnelle, 90, 133
- Erreur du PMC, 52
- Facteur taille, 45
- Fischer, 179
- Fonction
  - d'activation, 51
  - noyau, 55, 56
- Fréquence
  - conjointe, 83
  - marginale, 84
  - relative, 83
- Hotelling, 7
- Huygens, 184
- Hyper-tableau de contingence, 137
- Individu, 8
  - actif, 31
  - supplémentaire, 31, 158
- Inertie, 11, 13
  - expliquée, 29
  - taux, 30
- Interprétation
  - ACM, 153
  - ACP, 27
  - AFC, 105
  - règles, 34, 112, 161
- Invariance à l'échelle, 217
- Kernel*
  - Fischer Discriminant*, 214
  - Principal Component Analysis*, 56
- KFD, 214
- KPCA, 56
- Lagrangien, 12, 15, 47, 187
- Loi
  - du Chi-2, 197
  - normale multidimensionnelle, 30, 197
- Machines à vecteurs de support, 217
- Mahalanobis, 195
- Matrice
  - d'analyse discriminante, 208
  - des corrélations empiriques, 24
  - des covariances empiriques, 23
  - des covariances interclasses, 183
  - pondérées, 211
  - des covariances intra-classes, 183
  - des covariances totales, 183
  - mal conditionnée, 191

- Modalité, 81
- Moyenne empirique, 23
- Multiplicateurs de Lagrange, 12, 15, 47, 206
- Neurone, 51
- Noyau
  - conditionnellement défini positif, 59, 217
  - défini positif, 56, 214
  - gaussien, 56
  - polynomial, 56
  - puissance, 56
- Nuage
  - centré, 23
  - des individus, 10
  - des variables, 17
  - non centré, 114
- Observation, 8
- Pearson, 7
- Perceptrons multicouches, 50, 208
- PMC, 50, 208
- Poids synaptique, 51
- Points
  - individus, 8
  - variables, 8
- Pondération
  - des profils, 88
  - différenciée, 46, 192
- Pouvoir discriminant, 198
- Problème de valeurs propres
  - généralisé, 187, 223
  - ordinaire, 12, 17
- Procédure pas à pas, 199
- Profil
  - colonne, 87
  - modifié, 92
  - ligne, 85
  - modifié, 91
  - supplémentaire, 110
- Projection orthogonale, 10
- Pseudo-inverse Moore-Penrose, 209
- Qualité
  - d'approximation, 30
  - de représentation, 28, 107, 154
- Rangs, 49
  - corrigés, 49
- Rapport de corrélation, 149
- Reconstitution, 20
  - approximative, 20
  - exacte, 20
- Régression linéaire, 194
- Régularisation, 192
- Relation de Huygens, 184
- Relations
  - de transition, 19, 102
  - quasi-barycentriques, 102
- Représentation simultanée, 102, 149
- Réseaux de neurones, 50
- Robustesse, 49
- Sebestyen, 202
- Seuil, 51
- Support Vector Machines*, 218
- SVM, 218
- Tableau
  - de Burt, 137
  - de contingences, 82
  - de notes, 127
  - disjonctif complet, 136
- TdB, 137
- TDC, 136
- Test
  - de Kullback, 198
  - d'indépendance, 105
  - du Chi-2, 105
  - du Lambda de Wilks, 197
- Transformation affine, 51
- Valeur
  - d'activation, 51
  - propre, 13
- Valeur-test, 160



x

Validation croisée, 31

Variable

active, 31, 110

canonique, 177

nominale, 33, 81

quantitative, 8, 103

supplémentaire, 31, 110, 158, 173

Variance empirique, 24

Vecteur

moyenne, 22

propre, 13