absence of cluster structure 128, 129, 262–264
acoustic confusion matrix, application to 223
activation function 250
adaptive procedures 123
ADCLUS 223, 225, 231
additive clustering 223–226
additive inequality 89
additive tree 89, 91, 231, 239
adjusted cluster recovery 271–272
adjusted Rand index 265, 267, 272, 284, 287, 370
admissibility properties 80, 93–94
adolescent antisocial behaviour, application to 208–211
aesthetic judgement of painters, application to 136–141
age of onset of schizophrenia, application to 166–171
agglomerative methods 73–84, 98–101
centroid linkage clustering 75–76, 79, 83, 92, 241
complete linkage clustering 61, 76, 79, 81–83, 98, 266, 287–288
median linkage clustering 78–80, 83
single linkage clustering 61, 73–75, 76, 79, 81, 91, 96, 216, 219,
weighted average linkage clustering 77–78, 79
agglomerative methods, application empirical studies 83–84
illustration of the general procedure 73–76
problems in 80–83
recurrence formula 266, 288
summary table 79
aggregation index (pyramids) 229
agnes, R function 260
AIC (see Akaike’s information criterion) 160–161, 201
air pollution for US cities, application to 25–29, 133–136
Akaike’s information criterion 160–161, 201
algorithms, hierarchical 96
ALT (see autoregressive latent trajectory models)
alternating expectation-conditional maximization algorithm 197
angular separation 51
applications of cluster analysis adolescent antisocial behaviour 211–212
aesthetic judgement of painters 218
age of onset of schizophrenia 166–171
air pollution for US cities 25–29, 133–136
appropriateness of behaviour 221–222
archaeology 12, 83, 222, 246–249
astronomy 2–3, 9–10, 16–18, 21, 23
bioinformatics and genetics 12–13
body measurements 24–26
breast cancer 254–255
chemical compounds 285–287
clustering gene tests 171–173
INDEX 323

bootstrap probability 158–159, 268
branch and bound algorithms 121
brand switching 237
breakdown 151–152
breast cancer, application to 242–243
British Household Panel Survey 105
bundle 233
Calinski and Harabasz index 127, 129, 133, 260
Calinski-Harabasz stopping rule 260
CAMAN, R function 166, 171
Canberra distance 50
CART (see classification and regression tree)
case weight 94
centroid (see also exemplar) 61, 76, 78, 97, 113, 122, 126, 276
centroid effect ratio 236
centroid linkage clustering 75–76, 79, 83, 92, 241
centrotype (see exemplar)
CER (see centroid effect ratio)
CFA (see confirmatory factor analysis)
chaining 79, 80, 92, 216
chemical compounds, application to 281–285
city block distance 50, 98, 245
classification 1–13
 numerical methods 4–7
 reasons for 3–4
classification and regression tree 85
classification likelihood estimation 147–150
CLIQUE 221, 223
clumping 222–223
clustan 260
clustangraphics 241, 260
cluster analysis
 choice of method 258–259
 definition 7–9
 examples of use 9–13
 typical steps 258, 261
cluster recovery 163, 271
external validation 285–287
influence of individual points 267, 272
split sample validation 270

stability 83, 126, 151, 269
uncertainty in individual clusters 268
cluster solutions, comparing 157–163, 264–267, 286–288
dendrograms 265–267
partitions 264–265
proximity matrices 267
cluster validity profile 268
cluster, geometrical interpretation 79
cluster, R package 56, 126, 128
cluster, spherical 79
clustering criteria 111, 143, 148–149
derived from continuous data 113–115
derived from dissimilarity matrix 112–113
properties of 115–121
clustering gene tests, application to 171–173
clusterRepro, R package 268
clusterSim, R package 56, 130
clustvarsels, R package 166, 176–177, 207–208
co-absences 46
cognitive psychology, application to 37–38
Cohen’s plot 268
cohesion function 222
complete linkage clustering 61, 76, 79, 81–83, 98, 266, 287–288
composition of mammals’ milk, application to 107–110
compound symmetry 191
concordance index 128
conditional independence 152–153, 200
confirmatory factor analysis 57, 189, 194–195, 203, 205–206
consensus clustering 271
consensus graph 284
consensus network 280
consensus tree 271, 279
constrained clustering 105, 237–242, 253, 280
constrained single linkage 241
contiguity constraints 240–242
cophenetic matrix 91, 226, 267, 279–281
correlation coefficient 50–52
correspondence analysis 273
covariance matrix 144, 146–148, 151, 154, 166, 174, 187–202
crime rates, application to 43–45
crisp methods 241
crosscorrelation (spectrograms) 98
CURE 97
daisy, R function 56
data mining 3
data warehouse 222
DBSCAN (see density based spatial clustering)
Delaunay triangulation 240
dendrogram 72, 75, 81, 88–89, 95, 106, 107, 109, 220, 226, 236, 260
comparing 91–92
measuring distortion 91–92
optimizing 89
seriation 107, 109
terminology 88–89
density estimation 19–24, 41
density reachable 221
density search methods 216–220
density-based spatial clustering 220–222, 240
det(W)
cluster criterion 115, 128, 139, 142, 148–149
elliptical clusters 116, 148–149
scale independence 116
similar shape problem 116–118
similar size problem 116–118
diabetes, application to 173–180
diagnostic key 85
diameter of cluster 112
DIANA 285–287
diana, R function 86, 286
different component densities (see mixed-mode data)
direct clustering (see clustering two-mode)
direct optimizing algorithm 96
discrimination 7
dissection, of a cluster 8
dissimilarity
angular separation 51
application 130
between clusters 77
definition 5, 43
furthest neighbour 61
Jukes-Cantor dissimilarity 48
measures for continuous data 49–53
distance
Canberra distance 50
city block distance 50
definition 5, 49
edit distance 60
Euclidean distance 49, 114–115
furthest-neighbour distance 61
genetic distance 90
Levenshtein distance 59
Mahalanobis generalized distance 62, 63, 64, 123
measures for continuous data 49–53
Minkowski distance 50
nearest-neighbour distance 61
divisive methods 73, 84–88, 105–107, 241
association analysis 84, 101
DIANA 86, 105, 106, 285–287
McNaughton-Smith procedure 86, 105, 106
monothetic 84–85
Piccarretta and Billari method 85–86, 105, 107
polythetic 86–88, 101–105
dolphin whistles, application to 98–101
double permutation test 266, 279
Duda and Hart index 95, 127, 129, 260
Dunn’s partition coefficient 246
dynamic graph 15
dynamic programming 121
dynamic tree cutting 95
dynamicTreeCut, R package 95
diabetes, application to 43–45

eigenvalue decomposition 148–150, 155, 166, 174
EMMIX 197
EMMIX WIRE 205, 206
employment states 58–59
entropy 201, 210, 211
Epanechnikov kernel 20
epidemiology 267
error variance technique 234–239
Espalier 89
Euclidean distance 49, 75, 114–115
Euclidean property 52
evolutionary studies, application to 91
exemplar (see also centroid) 86, 88, 89, 108, 109, 113, 122, 130, 241, 261
exploratory factor analysis 57, 189, 194, 207
expectation-maximization (EM)
  algorithm 145–150
convergence problems 150–151
singularities and degenerate
distribution 150–151, 158, 161, 163
  stochastic EM 150
factor (see latent variable)
factor analysis 192
factor loadings 192, 193, 194
factor model 57, 153, 188–189, 197, 201
factor structure 194–196
FANNY 245, 285–287
fanny, R function 245, 286
feed-forward neuronal network 250
finite mixtures 97, 242, 143–212, 281, 283, 286, 287,
  applications for Bernoulli (latent class)
    models 177–183
  applications for mixed mode data
    models 178–185
application for multivariate normal
  models 173–180
applications for structured data 202–212
application for univariate Gaussian
  models 166–174
dimension reduction (variable selection)
in 163–164
estimation, Bayesian 146, 151, 154–157
estimation, classification
  likelihood 147–150
estimation, maximum
  likelihood 145–150, 153, 154
for categorical data (latent classes) 152–153, 164,
  for different component
    densities 153–154
  for mixed-mode data 153–154
  for multivariate Bernoulli
    densities 152–153
  for multivariate normal
    distributions 146–150
  for multivariate t-distributions 151–152
  for regression models 165
  for structured data 190–192
  of factor models 192–197
  of longitudinal models 197–202
  with unknown number of components and
    model structure 157–163
flexclust, R package 126
four point condition 89
functional magnetic resonance imaging,
  application to 18–19, 25, 27–29, 178–185
furthest neighbour distance 61, 76, 79
fuzzy clustering 242–249
fuzzy c-means, see fuzzy k-means
fuzzy DBSCAN 245
fuzzy k-means 245
fuzzy set theory 244
GAP-statistic 129, 130
Gaussian kernel 20
generalized linear model 165
generalized autoregressive parameters 201
gene sequences of yeasts, application
to 230–231
genetic algorithm 123
genetic data sets, application to 285–287
genetic distance 90
Genstat 260
gllamm 166, 197
global optima 122
globalization of cities, application
to 101–102, 104, 106
GMM (see growth mixture model)
GOM (see grade of membership)
Goodman and Kruskal’s γ 91, 265
Gower’s general similarity 54, 102, 130
grade of membership 234, 245
graphical display, of clusters 16–41, 273–279
group average linkage clustering (see
  average linkage clustering)
growth curve model 197–198, 201
growth factors 197, 198, 201, 210
growth mixture model 197, 198, 201, 209, 211
hclust, R package 260
heterogeneity 112
heuristic criteria 119, 143
HICLAS (see hierarchical classes) 250
hidden layer 250
hierarchical algorithms 96–97
hierarchical classes 232–234
hierarchical methods (see agglomerative methods, divisive methods) 232–234
high-dimensional data 13, 152, 161, 163, 164, 196
hill-climbing algorithm 121–124
histograms 16
histology, application to 16–17
homogeneity, of a cluster 7, 84, 86, 112
Hoplites producta (bee forms), application to 217–218
ideal type (see exemplar) 99
IGP (see in-group proportion) 99
image processing, application to 237, 240
importance of a variable 63, 64
INDCLUS 225
individual points, influence 267, 271–274
infants’ medical characteristics, application to 178, 183
information content 84
information criteria 160–161
Akaike’s information criterion 160–161, 201
Bayesian information criterion 160–162, 164, 166–168, 171, 173–176, 201, 203, 206–208, 210, 211
Bayesian information criterion, sample size adjusted 161–162
in-group proportion 268
insecticide tolerance, application to 162
internal cluster quality 267
internal cohesiveness 268
internet, application to 237, 254–255
invariance
under relabeling 223
under scaling 116, 260
inversion (see reversal) 99
Italian wines, application to 207–208
Jaccard coefficient 47, 102, 233, 234, 279, 282
jackknife correlation 52
Jaro similarity measure 60
Jukes-Cantor dissimilarity 48, 215
kappa coefficient 264
kcca function 126
kernel density estimates 20–24
kernel function 20
keyframes 239
\(k\)-means 98, 122, 123, 124–126, 133, 150, 215, 253, 260, 267, 274, 279, 285, 287
\(k\)-medians 122, 126, 260
Kohonen self organising map 215, 252–254, 273, 286
label switching
Bayesian estimation 156
during bootstrapping 159
Lance and Williams flexible method 78
parameters 80
recurrence formula 78, 94
large data sets 97–98, 253
latent class analysis 97, 144, 152–153, 164, 177–78, 181, 200, 245
latent class growth analysis 200
latent profile analysis 203
latent variable 144, 154, 155, 187, 189, 192, 193, 196, 197, 198
LatentGOLD 166, 197
layered feed forward neural network 250
LCGA (see latent class growth analysis)
Levenshtein distance 59
librarianship, application to 72
life histories, application to 85, 105–106
linguistics, application to 279–280
log-likelihood ratio test (LRT) 157–160
linkage parameter 217
LMR LRT (see Lo-Mendell-Rubin likelihood ratio test)
local optimum/maximum 122, 150
local independence (see also conditional independence) 153
Lo-Mendell-Rubin likelihood ratio test 158–159, 201, 210
longitudinal data 188, 197–199, 209
Mahalanobis generalized distance measure 62, 63, 64, 123
mammals’ milk, application to 107–110
Manhattan distance 50
Mantel test 266, 279
MAP (see maximum a posteriori probability)
MAPCLUS 225
market research, application to 9, 94
Markov chain Monte Carlo (MCMC) methods 146, 154–156, 157, 162, 162–163, 164
Marriott’s procedure 128, 139
masking variable 261
matching coefficient 47
matrix reordering 232
maximal complete subgraph 223, 225
maximal connected set 218
maximum a posteriori probability 145, 192, 201, 206, 207, 210, 211
maximum a posteriori estimate 151
maximum likelihood estimation 146–150, 153, 154, 192
McCulloch Pitts neuron 250
MCLUST family 149–150, 155, 201, 207
mclust, R package 165–167, 177, 208, 260
MCMC (see Markov chain Monte Carlo)
McNaughton-Smith procedure 86
MDS (see multidimensional scaling)
median linkage clustering 76, 79
medicine, application to 85, 241
medieval tiles, application to 273
medoid (see also exemplar) 86, 89, 105, 113, 122, 130, 133
membership function 241
meta analysis of medical data, application to 165
meteorology, applications to 11–12, 98
metric inequality 49
metric scaling (see multidimensional scaling)
minimum spanning tree 80, 96
Minkowski distance 50
missing value 85, 152, 154, 230, 261
mixed-mode data 54, 144, 153–154, 178–185
mixture item response theory modelling 197
mixture latent trait modelling 197
mixture of factor analysers model 152, 164, 195, 196, 197, 202–203
mixtures model (See finite mixture) mode analysis 216–217
Mojena’s rule 95–96
mona, R function 260
monothetic divisive methods 84–86, 101
monothetic system 2
monotone admissibility 69, 94
Mplus 166, 197, 201, 203, 210, 211
multidimensional scaling 36–38, 41, 226, 253, 283–284
multilayer perceptron 250
multimedia 239
museology, application to 72
mutual information distance 52
natural language 244
nearest neighbour clustering (see single linkage clustering)
nearest neighbour methods 217–220, 287
nearest-neighbour distance 61, 73, 79
Needleman-Wunsch algorithm 60
needs of psychiatric patients, application to 101–103
neighbourhood graph 274
Netherlands Twin Registry, application to 202–205
neuron 250
noise 83, 150, 221
non-specific back pain, application to 141–142
non-uniqueness 83
normal information radius 62
number of clusters 95–96, 126–130, 143, 157–163, 168–173, 261
Beale’s F-test 95, 127, 130
best cut 95
Calinski and Harabasz index 127, 129, 133, 260
concordance index 128
Duda and Hart index 95, 127, 129, 260
GAP-statistics 129, 130
Marriott’s procedure 128, 139
Mojena’s rule 95–96
upper tail rule 95, 108
silhouette plot 128–129, 130, 246, 247, 249, 268, 273, 274
number of clusters, comparison of procedures 158–159
number of groups (see number of clusters)
observational errors 83
oceanography, application to 221–223
OMA (see optimal matching algorithm)
OMA distance 86
optimal matching algorithms 60, 86, 215
optimization algorithms 121–126, 242
alternating expectation-conditional maximization algorithm 197
branch and bound algorithms 121
classification likelihood 147–148
dynamic programming 121, 241
expectation-maximization (EM) algorithm 145–150, 153, 154
genetic algorithm 123
hill climbing algorithm 121–124
k-means 98, 122, 123, 124–126, 133, 215, 260, 275, 286
k-medians 122, 126, 260
simulated annealing 123
 taboo search algorithm 123
variable neighbourhood search 123
optimization clustering 111–142
outlier 79, 81, 97, 150, 151, 221, 246
overlapping clusters 222–231
PAM (see partitioning around medoids)
pam, R function 126, 130, 260
parameter reduction techniques 163–164, 177, 196
parsimonious covariance structure 148–150, 195–196, 201–202
parsimonious Gaussian mixture models 148–150, 196, 207–208
parsimonious tree 96
particle physics experiment, application to 35–36
partitioning around medoids 122, 130
partitioning methods (see optimization methods)
path length tree (see additive tree)
PCA (see principal components analysis)
Pearson’s crabs, application to 145
perinatal depression, application to 211–212
permutation test 266, 281
PHYLIP 90
Piccarreta and Billari method 85–86, 105, 107
polythetic divisive method 86–88, 101–105
polythetic system 2
posterior distribution (Bayesian estimation) 155–157, 162
posterior probability 144–145, 146–148, 151, 153
precision 255
principal components analysis 29–32, 163, 192, 195, 246, 274
prior distribution 154–155, 155–156, 157, 162
projection index 33
projection pursuit 32–36, 274
projection, of multivariate data 29–38
proximity 43–69
choice of 68–69
definition 5, 43
inter-group measures 61–63
measures for structured data 56–60
proxy, R package 56
pruning 86
psychiatry, application to 10–11, 101, 110–111, 234–235
pyramids 89, 226–231
Q analysis 5
quantile-quantile plot 269
Rand index 261, 264–265, 284
random coefficients (see latent variable)
random effects 165
random graph hypothesis 263
random position hypothesis 263
random tree model 266
recall 255
rectangular kernel 20
reference vector 56–57, 188–189, 191, 198, 199
repeated measures 5, 56, 188
replicated microarray experiments, application to 205–207
replicated observations 94
reversal 79, 83, 92, 93, 241
robustness 267
of model selection criteria 161
rock crabs, application to 34–35, 152, 164
Roman glass composition, application to 246–249
Romano-British pottery, application to 30–33, 39–41, 275–276, 279
sample size adjusted Bayesian information criteria (see Bayesian information criteria)
SAS 56, 126
scaling 222–223
scatterplot matrix 24–29
scatterplots 16–19
self organising map (see Kohonen self organising map)
SEM (see structural equation model)
semi-supervised clustering method 237
SEMM (see structural equation mixture model)
separation, between clusters 7, 112, 161
sequence analysis 59
sequence data 85, 105
sequences 58–59, 85–86, 105
seriation 226, 228, 241
shadow value 274
significance test 261
silhouette index (silhouette value) 246, 268
silhouette plot 128–129, 130, 246, 247, 249, 268, 273, 274
similar shape problem 116–118
similar size problem 116–118
similarity
  correlation coefficient 50–52
definition 5, 46
Gower’s general similarity 54–56, 102, 130
Jaccard coefficient 47, 101, 233, 234, 280
Jaro similarity measure 60
matching coefficient 47
measures for binary data 46–47
measures for categorical data with more than two levels 47–49
measures for data containing both continuous and categorical variables 54–56
simulated annealing 123
single linkage clustering 61, 73–75, 76, 79, 80, 81, 91, 96, 216, 219, 266, 288
single unit perceptron 250
smoothing parameter 20
social relations in monastery, application to 225–226, 228
social systems, application to 72
software 35, 40, 86, 126, 165–166, 260
BMDP 260
Clustan 260
Clustangraphics 241, 260
EMMIX 197
EMMIX WIRE 205, 206
Genstat 260
LatentGOLD 166, 197
Mplus 166, 197, 201, 203, 210, 211
PHYLIP 90
SAS 56, 126
SPSS 56, 97, 126, 260
S-Plus 86, 260
Stata 56, 60, 126, 129, 166, 197, 260
Statistica 260
SOM (see Kohonen self organising map)
space conserving method 92
space contraction 92
space dilation 92
spectral clustering 287
spectrogram 98
spherical structure 79, 116
split sample validation 269–271
S-Plus 86, 260
SPSS 56, 97, 126, 260
SPSS TwoStep, SPSS component 97
sqom, Stata command 60
stability of clustering (see cluster solutions, assessment of quality)
standardization 64, 67–68, 97, 108, 115, 123, 261
star index 112
Stata 56, 60, 126, 129, 166, 197, 260
state permanence sequence 86, 105
statistica 260
INDEX 329
statistical models, for cluster analysis (see finite mixture models, model-based cluster analysis) 119, 143–186, 187–213
stochastic expectation-maximization algorithm 150
strength of membership 241
stripes plot 276–278, 280
structural equation mixture model 155, 166, 197, 201, 210, 213
structural equation model 197, 202, 213
structured data
definition 5, 56
model-based cluster analysis for 144, 187–213
subjective rating 94
sum of squares method 78
sum-of-squares 77, 78, 114
sums of the stars criterion 122
supervised learning 7
Swadesh list 48, 280
t-factor analyzer mixture model 152
tabu search algorithm 123
taxicab distance 50
taxmap method 216
taxonomy 2, 5
teaching behaviour, application
to 177–278, 181–182
text, clustering of 215, 254
three-dimensional graph 38–41
ties 83
time space clustering 221, 237, 267
trace($\text{BW}^{-1}$)
cluster criterion 115
scale independence 116
trace($W$)
algorithm for minimization 122, 123
cluster criterion 114–115, 141, 149
scale dependence 115, 123
similar shape problem 116
similar size problem 116–118
spherical structure 116, 149
training, of neural networks 252
tree fitting methods 213, 239
tree topology 88, 266
trellis graph 38–41
triangular kernel 20
two-mode clustering 231–237
two-way clustering/joining (see two-mode clustering)
ultrametric property 65, 92, 281
ultrametric tree 231, 239
uniformity hypothesis 263
unimodal null hypothesis 263
unique variance 192, 193, 194
univariate plot, of clusters 16–29
unsupervised pattern recognition 5
UPGMA 76, 79, 286
UPGMC 76, 79
upper tail rule 95, 108
variable neighbourhood search 123
variable selection 66, 163–164, 166, 176–177
video games, application to 54–56, 130–133
Voronoi diagram 240
weighted average linkage clustering 78, 79
weighting variables 63–67, 261
whisky tasting, application to 279–281
winning neuron 252–253
within group sum-of-squares 114
women’s life histories, application
to 105–107
Wong and Lane’s method 218–220
WPGMA 78–79
WPGMC 76, 79
z-scoring (see standardization)
z-values, p values transformation in 170–171, 174