

Index

- Abell, G. O., galaxies, 6, 25
- Adding technique, algorithm, 222
definition, 12
- Amalgamation rules, 216
- Analysis of variance, for evaluating clusters, 13
for K-means clusters, 89
variance components, 332
- Ant, 2
classification of, 16
eyes of, 9
- Anthropology, 5, 24
- Archaeology, 5, 24
- Automatic interaction detection, 337
- Average linkage, on bacteria, 20
on firms, 27
as a joining algorithm, 222, 227
on skulls, 5, 24
on stone tools, 24
superior, 4
- Bacteria, 4, 21
- Bailey, D. C., clustering variables and cases, 251, 269
factor analysis, 326
profiles, 40
- Barlow, R. E., monotone fitting, 204, 212
- Baron, D. N., classification of liver disease, 17
- Bartholomew, D. J., monotone fitting, 204, 212
- Barton, W. E., classification in psychiatry, 23
- Bassett, I. J., classification of pollen, 3, 17
- Bayes techniques, in K-means, 93
in mixture model, 115, 123
- Bellman, R. E., dynamic programming, 132, 140
- Bertin, J., beautiful graphs, 164
drawing trees, 162
- Bilines, need work, 39
- Bimodality, measure of, 99
test for, 98
- Blackwelder, R. A., taxonomy, 1, 14
weighing necessary, 67
- Block clusters, 288, 324, 342
definition, 11
histogram, 40
- Blood, 4
- Boorman, K. E., blood group serology, 4, 17
- Borradaile, L. A., classification of ants, 2, 16
- Bostonians, stealing cars, 33
- Bouckaert, A., goiters, 4, 18
- Boundaries of clusters, 191
- Boyce, A. J., skulls, 5, 24
- Bremner, J. M., monotone fitting, 204, 212
- Brunk, H. D., monotone fitting, 204, 212
- Burbank, F., epidemiology of cancer, 4, 18
- Camin, J. H., evolutionary trees, 233, 245
- Cases, 9
- Category variables, distances for, 64
- Cattell, R. B., coefficient of pattern similarity, 67
- Cavalli-Sforza, L. L., evolutionary tree models, 245
- Chakraverty, P., viruses, 4, 18
- Chernoff, H., faces displaying shellfish, 40
- Cirrhosis, 3, 4
- Clark, P. J., packing distances, 5, 25
- Clustering, diameters, 157
estimating missing values, 9
functions of, 6
interpretation, 12
multivariate histograms, 9
names for, 1, 156
stability, 14
- Cole, J. O., classification in psychiatry, 23
- Cole, J. W. L., linguistics, 26
- Complete linkage, on goiters, 18
subjective partitions, 27

- Contiguity, 156
- Convexity, joining algorithm, 212
of K-means clusters, 94
- Cormack, R. M., discussion after, 8
survey paper, 1
waste of time, 15
- Correlation, as euclidean distance, 64, 191
- Cox, D. R., optimal classification of a normal distribution, 95, 105
- Crimes in U.S. cities, 28, 342
linearly optimal profiles, 36
rescaled data, 32
- Crompton, C. W., classification of pollen, 3, 17
- Cronbach, L. J., assessing similarity, 67
Mahalanobis distance, 63
- Dalenius, T., optimal stratification, 95, 106
- Darlington, P. J., 3
abandonment of reality, 17
- Darwin's evolutionary theory, 2
- Data structures, 9
- Dendrograms, 164
- Dentition of mammals, 170
- Diabetes mellitus, 3
- Dickerson, R. E., cytochrome-c, 240
- Discriminant analysis, for evaluating clusters, 13
in prediction, 331
separating K-means clusters, 94
- Disease, 3
- Distance, 9
matrix, 10
separated from clustering, 191
- Ditto algorithm, 143
- Dodd, B. E., blood group serology, 4, 17
- Doll, R., leukemia, 20
- Donnelly, W., leukemia, 21
- Dreyfus, S. E., dynamic programming, 140
- Dupont, P. F., classification of yeasts using DNA, 3, 17
- Dyen, I., linguistics, 5, 26
- Economics, 5
- Edwards, A. W. F., evolutionary tree models, 245
- Eigenvectors, 35
closest fitting plane, 39
interpreting Mahalanobis distance, 63
sorting, 175
splitting into two clusters, 99
- Electrocardiograms, 4, 21
- Engelman, L., test for clusters, 97, 98, 106
- Epidemiology, 4
- Erdos, P., random graphs, 202, 212
- Estabrooke, G. J., distances before clusters, 191
evolutionary trees, 233, 245
moats, 202, 212
- Euclidean distance, 58, 60
distributions of, 66
generalized, 63
plotted to detect clusters, 65
weighted, 60
- European languages, 243
- Evans, F. C., packing distances, 5, 25
- Everitt, B. S., classification in psychiatry, 5, 22
- Evolutionary tree construction, 237
- Faces, 39
- Factor analysis, algorithms, 313
in block clustering, 290, 324
on diabetes, 20
oblique factors, 320
in psychiatry, 5, 22
in psychology, 6
- Farris, J. S., link distance to ultrametrics, 162
Wagner trees, 246
- Feinstein, A. R., classification of disease, 3, 19
- Fisher, L., evaluation of K-means, 93, 106
- Fisher, W. D., clustering in economics, 5, 26, 268
- Fisher's algorithm, applied to Olympic data, 137
exact optimization, 130
time and space requirements, 137
- Fitch, W. M., minimum mutation fit, 233, 246
mutation distances, 209
- Flieger, W., expectation of life, 182
- Football prediction, 217
- Ford, T. L. E., butterflies, 150
- Frank, R. E., market research applications, 5, 26
- Fraser, P. M., classification of liver disease, 17
- Friedman, H. P., stages of critically ill, 19
weighting variables, 58
within-cluster covariance matrix, 63, 68
- Friedman-Rubin technique, in depression, 23
- Galaxies, 6
- Gallimaufry cluster, 147
- Gap, 191
- Genghis Khan, descendants, 164
- Gleser, G. C., assessing similarity, 67
Mahalanobis distance, 63
- Goiter, 4
- Goldwyn, R. M., stages of septic process, 3, 4, 19
- Good, I. J., 1
two-way splitting, 15, 268
- Goodfellow, N., bacteria, 4, 20
- Goodman, L. A., measures of association, 13, 15
- Goronzy, F., business enterprises, 5, 26
- Gourlay, A. J., classification in psychiatry, 22

- Gower, J. C., minimum spanning trees, 191, 201, 213
 spirited defense, 8
- Green, P. E., market research, 5, 26
- Hall, C. T., yeast, 279
- Hammersley, J. M., monotone fits, 312
- Harman, H. H., factor analysis, 325
- Hartigan, J. A., clusters given distances, 191
 direct splitting, 251
 minimum mutation fit, 233, 246
 optimal trees, 200, 213
 test for clusters, 97, 98, 106
 ultrametrics, 160, 164
- Hautaluoma, J., psychiatric syndromes, 5, 23
- Hayhoe, F. G. J., leukemia, 4, 20
- Heart disease, 4
- Hendrick, L. R., classification of yeast, using DNA, 3, 17
- Hess, S. W., legislative districts, 5, 27
- Hierarchical clustering, *see trees*
- Hodson, F. R., archaeological applications, 25
 tools, broaches, 5, 24
- Homology problem, 9, 15
- Huizinger, J., distances in anthropology, 5, 25
- Imms, A. D., insects cerci, 233, 246
- Information, association between two partitions, 13
- Insect, classified by different systems, 14
 wings, 10
- James, A. T., linguistics, 26
- Jardine, C. J., distances, 191
 optimally fitting trees, 200, 213
 ultrametrics, 164
- Jardine, N., mathematical taxonomy, 1, 15
 optimally fitting trees, 200, 213
 ultrametrics, 160, 164
- Jigsaw puzzle, 75
- Johnson, S. C., complete linkage, 200, 213
 single-linkage clusters, 191
 ultrametrics, 160, 164
- Joining, algorithms, 216
 definition, 11
 in factor analysis, 319
 probability models, 229
 two-way algorithms, 278
- K-means, algorithm, 84–112
 on axes, 25
 on broaches, 25
 on cities, 26
 convexity of clusters, 94
 global optimality, 103
 in psychiatry, 22
 variations on, 102
- Kaiser, H. F., legislative districts, 5, 27
- Kant, I., classification of mental disorders, 4, 23
- Katz, M. M., classification in psychiatry, 4, 23
- Kendall, D. G., archaeology, 25
- Kendall, R. E., classification in psychiatry, 22
- Keyfitz, N., expectations of life, 182
- King, B. F., stocks, 5, 27
- Klett, C. J., psychiatric syndromes, 23
- Knospe, W. H., leukemia, 21
- Knusman, R., diabetes, 3, 20
- Kruskal, J. B., minimum spanning tree, 201, 213
- Kruskal, W. H., measures of association, 13, 15
- Lance, G. N., joining of algorithms, 200, 213, 222, 230
- Landsteiner, 4
- Leader algorithm, 75
- Legislative districts, 6, 27
- Leukemia, 21, 334
- Levelt, W. J. M., triads, 184
- Lieth, H., phytosociology, 5, 26
- Likelihood, generalized K-means, 93
 in mixture models, 113–125
- Ling, R., reality of clusters, 202, 213
- Linguistics, 5
- Linnaeus, 1
- Livermore, T. L., Civil War casualties, 121
- Local optimization, 84
- Lockwood, drawing trees, 162, 164
- Lorr, M., psychiatric syndromes, 5, 23
- MacQueen, J., K-means algorithm, 106
- Mahalanobis, P. C., 63, 68
- Mahalanobis distance, 16
 bad effects in weighting, 63, 67, 68
- Mammals' milk data, 6
- Manic depressive, unobservable, 4
- Manning, R. T., heart disease, 4, 20
- Manual of disease classification, 3, 20
- Margoliash, mutation distances, 209
- Market research, 5, 26
- Maurer, W., profile algorithm, 32
- McNair, D. M., psychiatric syndromes, 23
- McQuitty, L. L., maximum joining algorithm, 213
- Miller, G. A., subjective partitions, 6, 27
- Minimum mutation fit, 233
 joining algorithm for, 237
 probability theory for, 236
 real variables, 244
- Minimum spanning tree, 201
- Minkowski distances, 65
- Minor planets, grouped sightings table, 2
 keeping track of, 1
- Missing values, 267
- Mixture model, 13
 algorithms, 113–129

- related to K-means, 124
- Modes, as cluster centres, 143
 - density contour clusters, 211
 - by maximum likelihood, 140
- related to single linkage, 15
- Moore, C. W., phytosociology, 5, 26
- Moore, P., moons and planets, 122
- Morgan, AID technique, 331
- Multivariate normal, 116

- Nosology, 1, 3
- Numerical taxonomy, 1
 - examples of, 3
 - logically unsound, 8

- Okajima, M., electrocardiograms, 21
- Overlapping clusters, 15, 16
 - for broaches, 25

- Palmer, E. L., dentition of mammals, 170
- Papageorge, C., yeast, 279
- Partition, definition, 11
 - by exact optimization, 130
 - models, 11
 - neighborhoods, 84
 - optimal, 95
 - quick algorithms, 74
- Paykel, E. S., classifications of depressed patients, 5, 23
- Pearson, K., coefficient of racial likeness, 68
- Peterson, S., presidential elections, 252
- Phytosociology, 5
- Pollen, 3, 17
- Potts, F. A., classification of ants, 2, 16
- Prediction, 330
- Prevot, A. H., bacteria, 4, 21
- Probability of rain, 9
- Profiles, 28-57
 - algorithm, 29
 - linearly optimal, 34
 - rank profiles, 32
 - transposed, 39
- Psychiatry, 4

- Quaglino, D., leukemia, 20

- Range, 268, 291
- Rao, C. R., Indian caste measurements, 324
 - Mahalanobis distance, 63, 68
- Ratio measure for validity of partition, 91
- Regression, 323, 330
- Renal disease, 4
- Renyi, A., random graphs, 202, 212
- Root of a matrix, 314
- Ross, G. J. S., single-linkage clusters, 191, 201, 213
- Rubin, J., weighting variables, 58

- within-cluster covariance matrix, 63, 68

- Scaling, 299
- Scammon, R. M., Connecticut votes, 267
- Schizophrenic, 5
- Schrek, R., leukemia, 4, 21
- Sebestyen, G. S., K-means type algorithm, 107
- Severity of renal disease, 18
- Shakow, D., variability of psychiatric classification, 4, 24
- Sibson, R., mathematical taxonomy, 1, 15
 - optimally fitting trees, 200, 213
 - ultrametrics, 160, 164
- Siegel, J. H., stages of septic process, 19
- Similarities, 9
- Simple structure, 314
- Single linkage, on airline distances, 195
 - algorithm, 191
 - on bacteria, 20
 - beaten, 4
 - broaches, 25
 - on cirrhosis, 17, 22
 - continuous function of dissimilarities, 16
 - density contour tree, 205
 - exact optimization, 199, 200
 - on goiters, 18
 - invariance under transformation, 195
 - on leukemia, 20
 - partitions, 199
 - reducing distance calculations, 212
 - sausage clusters, 200
 - significance tests, 204
 - subjective partitions, 27
 - on viruses, 18
- Small, E., classification of pollen, 3, 17
- Sneth, P. H. A., don't weight, 68
 - joining, 200, 214
 - numerical taxonomy, 1, 3, 15
 - single-linkage algorithm, 191, 213
 - weights, 61
- Sokal, R. R., don't weight, 68
 - drawing trees, 162, 164
 - evolutionary trees, 213, 245
 - foundation of numerical taxonomy, 1, 3, 16
 - joining techniques, 200, 214
 - single-linkage clusters, 191
 - weights, 61
- Sonnquist, J. A., splitting technique, 251, 269, 331
- Sorting, algorithm, 78, 173
 - definition, 11
- Species, 1
- Spiral search algorithm, 196
- Splitting, definition, 11
 - direct splitting, 251
 - two-way splitting, 260
- Stark, L., 4

- electrocardiograms, 4, 21
- Stearn, W. T., 3, 17
- Struhsaker, T. T., vervet monkeys, 149
- Subsampling, in K-means, 105
- Switching, definition, 11

- Tantu, P., archaeology, 25
- Temkin, O., history of medical classification, 3, 21
- Threshold, in quick algorithm, 78
- Tied distances, 292
- Toeller, M., diabetes, 3, 20
- Tree, complete search, 185
 - definition, 11, 155
 - as a directed graph, 158
 - drawings, 155
 - leader algorithm, 169
 - mixture model, 124
 - models, 11
 - natural trees, 164
 - as triads, 177
- Triads, algorithm using, 177
 - probability model, 187
 - similarity between trees, 13
 - triads-leader algorithm, 181
- Tryon, R. C., clustering variables then cases, 251, 269
 - factor analysis, 326
 - profiles, 40
- Tumors, 4
- Two-way clustering, 10
- Tygstrup, N., cirrhosis, 3, 22

- Ultrametric, 11, 16
 - definition, 160

- Van Ness, J. W., evolution of K-means, 93
- Variables, 9
- Variance-components algorithm, 332
- Vietnam combat deaths, 139
- Viruses, 4, 21, 22
- Voting, Congress, 242
 - Connecticut, 267
 - presidential elections, 251
 - United Nations, 300

- Watson, L., heart disease, 4, 20
- Weaver, J. B., legislative districts, 5, 27
- Webb, C. D., yeast, 279
- Weights, 60
 - circularity using within-cluster variances, 62
 - for food data, 84, 91
 - inverse variances risky, 61
 - Mahalanobis distance dangerous, 63
 - in two-way joining, 292
- Weiner, J. S., distances in anthropology, 5, 25
- Whipple, G. M., electrocardiograms, 21
- Whittaker, R. H., phytosociology, 5, 26
- Wife, notices husband, 1
- Wildy, P., viruses, 4, 21
- Wiley, D. E., latent partition analysis, 6, 13, 27
- Williams, W. T., joining algorithms, 200, 213
- Wilner, B. I., viruses, 4, 22
- Wines, evaluation, 144
- Winkel, P., cirrhosis, 3, 22
- Wishart, D., reducing chaining, 214

- Zubin, J., schizophrenia, 5, 24