REFERENCES

Højsgaard, S., Skjøth, F., and Thiesson, B. User’s guide to Bifrost. Aalborg University, Aalborg, Denmark.
REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


applications of expert systems, pages 83–204, Sydney. New South Wales Institute of Technology.


REFERENCES

error propagation. In Rumelhart, D. E. and McClelland, J. L., editors, Parallel Distributed

18(3):257–266.

Sammut, C. (1988). Experimental results from an evaluation of algorithms that learn to
control dynamic systems. In LAIRD, J., editor, Proceedings of the fifth international
Morgan Kaufmann.

Sammut, C. (1994). Recent progress with boxes. To appear in Machine Intelligence and
Inductive Learning, Vol. 1 (eds. Furukawa, K. and Muggleton, S. H., new series of
Machine Intelligence, ed. in chief D. Michie), Oxford: Oxford University Press.

In Proceedings of the Seventh International Machine Learning Conference, pages 170–
178, Austin, Texas. Morgan Kaufmann.

and Edwards, P., editors, Proceedings of the Ninth International Workshop on Machine
Learning, pages 385–393. Morgan Kaufmann.


R. S. Michalski, J. C. and Mitchell, T., editors, Machine Learning: An Artificial Intelli-


Schumann, M., Lehrbach, T., and Bahrs, P. (1992). Versuche zur Kreditwur-
digkeitsprognose mit kunstlichen Neuronalen Netzen. Universitat Gottingen.


classifiers. In IJCNN-90: proceedings of the international joint conference on neural

and two statistical estimators in a sparse and noisy environment. In IJCNN-90:
proceedings of the international joint conference on neural networks, pages 289–292,
Ann Arbor, MI. IEEE Neural Networks Council.


Index

\( AC^2 \), 12, 67, 123, 218, 263  
\( H(C) \), 117, 173  
\( H(C, X) \), 117  
\( H(X) \), 116  
\( M(C, X) \), 117  
\( \hat{H}(X) \), 174

Accuracy, 7, 8  
ACE, 46  
Algorithms  
  - function approximation, 230  
Algorithms  
  - instance-based, 230  
Algorithms  
  - symbolic learning, 230  
ALLOC80, 33, 214, 227, 263  
Alternating Conditional Expectation, 46  
Analysis of results, 176  
AOCDL, 56  
AQ, 56, 57, 74, 77  
Aq, 237  
AQ11, 50, 54  
Architectures, 86  
Assistant, 65  
Attribute coding, 124  
Attribute entropy, 174  
Attribute noise, 174  
Attribute reduction, 120  
Attribute types, 214  
Attribute vector, 17  
Attributes, 1

Australian credit dataset, 134  
Background knowledge, 11, 241  
Backprop, 12, 110, 221, 263  
Bayes minimum cost rule, 13  
Bayes Rule, 40  
Bayes rule, 13, 14, 16, 17, 27  
Bayes theorem, 15  
Bayes Tree, 263  
Bayes tree, 12, 73, 123, 220  
Bayes-tree, 41  
Bayesian evidence, 100  
Bayesian methods, 29  
Bayesian Networks, 41  
Bayesian regularisation  
  - Cascade correlation, 98  
  - behavioural cloning, 261  
Belgian Power I dataset, 121, 163  
Belgian Power II dataset, 164  
Bias, 120, 122  
BIFROST, 43  
Binary attributes, 25  
Binomial, 26  
Bootstrap, 107–109  
BOXES, 234, 250

C4.5, 12, 63–65, 79, 80, 209, 219, 264  
CAL5, 71, 72, 263  
Cal5, 12, 70, 219  
Canonical correlation, 114, 173  
Canonical discriminants, 114
Canonical variates, 20, 114
CART, 12, 63, 64, 68, 123, 126, 132, 218, 219, 225, 264
Cascade, 110
Cascade correlation, 12, 97, 98, 263
CASTLE, 45, 217, 226, 263
Categorical variables, 17
Causal network, 41
Causal Networks, 41
CHAID, 62
Chernobyl, 7
Chi-square test of independence, 119
Choice of variables, 11
Chromosome dataset, 142
Class, 8
Class definitions, 7
Class entropy, 173
Class probability tree, 73
Class probability trees, 61
Classes, 1
Classical discrimination techniques, 17
Classification, 1, 8
Classification rule, 6
Classification: definition, 6
CLS, 52
Clustering, 1, 6
CN2, 12, 56, 57, 73–75, 77, 79, 218, 220
Code vector, 101
Coding of categories, 122
Combination of attributes, 121
Combinations of variables, 11
Comparative trials, 110
Complex, 56
Comprehensibility, 7, 214
Concept, 228
Concept learning, 51
Concept Learning System, 52
Concept-recognisers, 62
Condensed nearest neighbour, 35
Conditional dependency, 53
Conjugate gradient, 91
Constructive algorithms, 88
Constructive algorithms
pruning, 96
container cranes, 258
controller design, 246
Corr abs, 173
Correlation, 113
Correspondence Analysis, 185
Cost datasets, 176, 183
Cost matrices, 214, 224
Cost matrix, 221
Costs, 225
Covariance, 113
Covariance matrix, 19, 21
Cover, 56
Covering algorithm, 237
Credit datasets, 7, 8, 122, 132–135, 176
Credit management dataset, 122, 132
Credit scoring, 132
Cross validation, 107–109
Cross-entropy, 89
Cut20 dataset, 121, 146, 181
Cut50 dataset, 121, 146, 181
DAG (Directed acyclic graph), 41
Data
soybean, 50–52
Dataset
Australian credit, 134
Dataset
Belgian Power I, 163
Dataset
Belgian Power II, 164, 174
Dataset
chromosomes, 142
Dataset
credit management, 132
Dataset
cut, 146
Dataset
diabetes, 157
Dataset
DNA, 158
Dataset
German credit dataset, 153
Dataset
hand-written digits, 135
Dataset
head injury, 149
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Directed acyclic graph (DAG), 41</th>
</tr>
</thead>
<tbody>
<tr>
<td>heart disease, 152</td>
<td>Discrim, 17, 121, 126, 173, 214, 225</td>
</tr>
<tr>
<td>image segmentation, 145</td>
<td>Discrimination, 6, 8</td>
</tr>
<tr>
<td>Karhunen-Loeve Digits, 137</td>
<td>Distance, 161</td>
</tr>
<tr>
<td>letter recognition, 140</td>
<td>Distribution-free methods, 16</td>
</tr>
<tr>
<td>machine faults, 165</td>
<td>DNA dataset, 23, 122, 124, 158, 161, 222, 226</td>
</tr>
<tr>
<td>satellite image, 143</td>
<td>domain knowledge, 255</td>
</tr>
<tr>
<td>shuttle control, 154</td>
<td>Dominators, 186</td>
</tr>
<tr>
<td>Technical, 174</td>
<td>EA, 111</td>
</tr>
<tr>
<td>technical, 161</td>
<td>ECG, 52, 227</td>
</tr>
<tr>
<td>tsetse fly distribution, 167</td>
<td>Edited nearest neighbour, 35</td>
</tr>
<tr>
<td>vehicle recognition, 138</td>
<td>EN.attr, 118</td>
</tr>
<tr>
<td>Credit management, 124</td>
<td>Entropy, 70, 76–78</td>
</tr>
<tr>
<td>cut, 181</td>
<td>Entropy estimation, 117</td>
</tr>
<tr>
<td>Karhunen-Loeve Digits, 193</td>
<td>Entropy of attributes, 116</td>
</tr>
<tr>
<td>shuttle control, 193</td>
<td>Entropy of classes, 117</td>
</tr>
<tr>
<td>Shuttle, 173</td>
<td>Epistemologically adequate, 80</td>
</tr>
<tr>
<td>Dataset characterisation, 112</td>
<td>Equivalent number of attributes, 118</td>
</tr>
<tr>
<td>Dataset collection, 124</td>
<td>Error rate, 194</td>
</tr>
<tr>
<td>Decision class, 14</td>
<td>Error rate estimation, 107</td>
</tr>
<tr>
<td>Decision problems, 1</td>
<td>Evaluation Assistant, 110, 111</td>
</tr>
<tr>
<td>Decision trees, 5, 9, 56, 73, 109, 121, 161, 217, 226</td>
<td>Examples of classifiers, 8</td>
</tr>
<tr>
<td>Default, 57, 80</td>
<td>Expert systems, 50</td>
</tr>
<tr>
<td>Default rule, 13</td>
<td>Extensions to linear discrimination, 12</td>
</tr>
<tr>
<td>Density estimates, 12</td>
<td>Features, 1</td>
</tr>
<tr>
<td>Density estimation, 30</td>
<td>Feed-forward networks, 96</td>
</tr>
<tr>
<td>Diabetes dataset, 157</td>
<td>Feedforward network, 88</td>
</tr>
<tr>
<td>Digits dataset, 135, 181, 223</td>
<td>First order logic, 230</td>
</tr>
<tr>
<td>DIPOL92, 12, 103, 223, 225, 263</td>
<td>Fisher’s linear discriminant, 9, 17</td>
</tr>
<tr>
<td></td>
<td>fractK, 170, 173</td>
</tr>
<tr>
<td></td>
<td>fractK, 114</td>
</tr>
<tr>
<td></td>
<td>Gaussian distribution, 20</td>
</tr>
<tr>
<td></td>
<td>General-to-specific, 54, 56, 57</td>
</tr>
<tr>
<td></td>
<td>Generalised Delta Rule, 86</td>
</tr>
<tr>
<td></td>
<td>Generalised linear models (GLM), 26</td>
</tr>
<tr>
<td></td>
<td>Genetic, 65</td>
</tr>
<tr>
<td></td>
<td>Genetic algorithms, 2, 5, 234</td>
</tr>
<tr>
<td></td>
<td>genetic algorithms, 252</td>
</tr>
<tr>
<td></td>
<td>German credit, 153</td>
</tr>
<tr>
<td></td>
<td>Gini function, 68</td>
</tr>
<tr>
<td></td>
<td>Gini index, 68</td>
</tr>
</tbody>
</table>
INDEX 287

GLIM, 26
GOLEM, 81
Golem, 244
Gradient descent, 90
Gradient descent
MLP, 92
Gradient descent
second-order, 91
Gradient methods, 93
Head dataset, 149, 173
head injury dataset, 23
Heart dataset, 152, 173
Heuristically adequate, 80
Hidden nodes, 109
Hierarchical clustering, 189, 192
Hierarchical structure, 2
Hierarchy, 120, 123
Human brain, 3
Hypothesis language, 54, 229
ID3, 160, 218, 219
ILP, 65, 81, 82
Image datasets, 176, 179, 182
Image segmentation, 112, 181
Impure, 60
Impure node, 57
Impurity, 57, 58, 60
IND Package, 40
IND package, 73
IndCART, 12, 219, 263
Indicator variables, 26
inductive learning, 254
Inductive Logic Programming, 81, 82
Inductive logic programming, 2, 5
Inductive Logic Programming (ILP), 50
Information measures, 116, 169
Information score, 203
Information theory, 116
Instance-based learning (IBL), 230
Iris data, 9
Irrelevant attributes, 119, 226
ISoft dataset, 123
ITrule, 12, 56, 57, 77, 78, 220, 265
J-measure, 56, 78
Jackknife, 32
Joint entropy, 117
K nearest neighbour, 160
K-Means clustering, 102
K-means clustering, 101
K-Nearest Neighbour, 35
K-Nearest neighbour, 10–12, 16, 126
k-Nearest neighbour, 29
k-NN, 160, 182, 216, 224, 227, 265
K-NN
   Cross validation, 36
K-R-K problem, 80–82
Kalman filter, 96
KARDIO, 52, 227
Kernel
   classifier, 33
Kernel
   window width, 33
Kernel density (ALLOC80), 12
Kernel density estimation, 30, 214
Kernel function, 31
Kernels, 32
KL digits dataset, 27, 121, 137, 170
Kohonen, 160, 222, 265
Kohonen networks, 85, 102
Kohonen self-organising net, 12
Kullback-Leibler information, 112
Kurtosis, 22, 115, 170
Layer
   hidden, 87
Layer
   input, 86
Layer
   output, 86
learning curves, 127
Learning graphical representations, 43
Learning vector quantization (LVQ), 12
Learning Vector Quantizer, 102
Learning vector quantizers, 102
Leave-one-out, 108
Letters dataset, 140, 208
Likelihood ratio, 27
Linear decision tree, 156
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear discriminant, 11, 12, 17, 104, 214</td>
<td></td>
</tr>
<tr>
<td>Linear discrimination, 27</td>
<td></td>
</tr>
<tr>
<td>Linear independent, 121</td>
<td></td>
</tr>
<tr>
<td>Linear regression, 26, 104</td>
<td></td>
</tr>
<tr>
<td>Linear threshold unit (LTU), 233</td>
<td></td>
</tr>
<tr>
<td>Linear transformation, 115</td>
<td></td>
</tr>
<tr>
<td>Linear trees, 56</td>
<td></td>
</tr>
<tr>
<td>Line searches, 91</td>
<td></td>
</tr>
<tr>
<td>Link function, 26</td>
<td></td>
</tr>
<tr>
<td>Log likelihood, 32</td>
<td></td>
</tr>
<tr>
<td>Logdisc, 24, 121, 263</td>
<td></td>
</tr>
<tr>
<td>Logistic discriminant, 17, 24</td>
<td></td>
</tr>
<tr>
<td>Logistic discrimination, 27</td>
<td></td>
</tr>
<tr>
<td>Logistic discrimination - programming, 25</td>
<td></td>
</tr>
<tr>
<td>LVQ, 102, 126, 221, 222, 264</td>
<td></td>
</tr>
<tr>
<td>M statistic, 113</td>
<td></td>
</tr>
<tr>
<td>Machine faults dataset, 165</td>
<td></td>
</tr>
<tr>
<td>Machine learning approaches, 16</td>
<td></td>
</tr>
<tr>
<td>Machine learning approaches to classifica-</td>
<td></td>
</tr>
<tr>
<td>tion, 2</td>
<td></td>
</tr>
<tr>
<td>MADALINE, 223</td>
<td></td>
</tr>
<tr>
<td>Manova, 20, 114, 173</td>
<td></td>
</tr>
<tr>
<td>Many categories, 121</td>
<td></td>
</tr>
<tr>
<td>Marginalisation, 98</td>
<td></td>
</tr>
<tr>
<td>MARS, 47</td>
<td></td>
</tr>
<tr>
<td>Maximum conditional likelihood, 25</td>
<td></td>
</tr>
<tr>
<td>Maximum likelihood, 20, 25, 32</td>
<td></td>
</tr>
<tr>
<td>McCulloch-Pitts neuron, 84</td>
<td></td>
</tr>
<tr>
<td>MDL, 80</td>
<td></td>
</tr>
<tr>
<td>Measure of collinearity, 114</td>
<td></td>
</tr>
<tr>
<td>Measures, 112, 209</td>
<td></td>
</tr>
<tr>
<td>Measures</td>
<td></td>
</tr>
<tr>
<td>Information-based, 169</td>
<td></td>
</tr>
<tr>
<td>Measures statistical, 169</td>
<td></td>
</tr>
<tr>
<td>Measures of normality, 114</td>
<td></td>
</tr>
<tr>
<td>Medical datasets, 217</td>
<td></td>
</tr>
<tr>
<td>Memory, 223</td>
<td></td>
</tr>
<tr>
<td>Mental fit, 50–52, 56, 79, 80</td>
<td></td>
</tr>
<tr>
<td>mental fit, 79</td>
<td></td>
</tr>
<tr>
<td>Metalevel learning, 197</td>
<td></td>
</tr>
<tr>
<td>Minimisation methods, 90</td>
<td></td>
</tr>
<tr>
<td>Minimum cost rule, 14</td>
<td></td>
</tr>
<tr>
<td>Minimum Description Length (MDL) Prin-</td>
<td></td>
</tr>
<tr>
<td>ciple, 80</td>
<td></td>
</tr>
<tr>
<td>Minimum risk rule, 14</td>
<td></td>
</tr>
<tr>
<td>Misclassification costs, 13, 14, 17, 58, 177</td>
<td></td>
</tr>
<tr>
<td>Missing values, 17, 66, 70, 76, 120, 214, 216</td>
<td></td>
</tr>
<tr>
<td>ML on ML, 211</td>
<td></td>
</tr>
<tr>
<td>MLP, 85–88</td>
<td></td>
</tr>
<tr>
<td>Mental fit, 51</td>
<td></td>
</tr>
<tr>
<td>Multi Layer Perceptron, 85–88</td>
<td></td>
</tr>
<tr>
<td>Multi Layer Perceptron functionality, 87</td>
<td></td>
</tr>
<tr>
<td>Multi-class trees, 58, 62</td>
<td></td>
</tr>
<tr>
<td>Multidimensional scaling, 187, 190</td>
<td></td>
</tr>
<tr>
<td>Multimodality, 112</td>
<td></td>
</tr>
<tr>
<td>Multivariate analysis of variance (Manova),</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Multivariate kurtosis, 115, 170</td>
<td></td>
</tr>
<tr>
<td>Multivariate normality, 114</td>
<td></td>
</tr>
<tr>
<td>Multivariate skewness, 115</td>
<td></td>
</tr>
<tr>
<td>Mutual information, 117, 119</td>
<td></td>
</tr>
<tr>
<td>Naive Bayes, 12, 40, 216, 263</td>
<td></td>
</tr>
<tr>
<td>Nearest neighbour, 7, 35</td>
<td></td>
</tr>
<tr>
<td>Nearest neighbour example, 36</td>
<td></td>
</tr>
<tr>
<td>Neural network approaches, 3, 16</td>
<td></td>
</tr>
<tr>
<td>Neural networks, 5, 221, 227</td>
<td></td>
</tr>
<tr>
<td>Neurons, 3</td>
<td></td>
</tr>
<tr>
<td>NewID, 12, 65, 66, 68, 122, 160, 218</td>
<td></td>
</tr>
<tr>
<td>No data rule, 13</td>
<td></td>
</tr>
<tr>
<td>Node hidden, 87</td>
<td></td>
</tr>
<tr>
<td>Node impure, 57</td>
<td></td>
</tr>
<tr>
<td>Node input, 87</td>
<td></td>
</tr>
<tr>
<td>Node output, 87</td>
<td></td>
</tr>
<tr>
<td>Node purity, 61</td>
<td></td>
</tr>
<tr>
<td>Node winning, 102</td>
<td></td>
</tr>
<tr>
<td>Noise, 56, 61, 73, 79, 216, 219, 223</td>
<td></td>
</tr>
<tr>
<td>Noise signal ratio, 119</td>
<td></td>
</tr>
</tbody>
</table>
Noisy, 57
Noisy data, 61
Nonlinear regression, 89
Nonparametric density estimator, 35
Nonparametric methods, 16, 29
Nonparametric statistics, 5
Normal distribution, 20
NS.ratio, 119, 174

Object recognition datasets, 180
Observation language, 53, 229
Odds, 25
Optimisation, 94
Ordered categories, 25
Over-fitting, 107
Overfitting, 63, 64

Parametric methods, 16
Partitioning as classification, 8
Parzen window, 30
Pattern recognition, 16
Perceptron, 86, 109, 232
Performance measures, 4
Performance prediction, 210
Plug-in estimates, 21
Polak-Ribiere, 92
pole balancing, 248
Polytrees, 43
Polytrees (CASTLE), 12
Polytrees
   as classifiers, 43
Pooled covariance matrix, 19
Prediction as classification, 8
Preprocessing, 120, 123
Primary attribute, 123
Prior
   uniform, 100
Prior probabilities, 13, 133
Probabilistic inference, 42
Products of attributes, 25
Projection pursuit, 37, 216
Projection pursuit (SMART), 12
Projection pursuit
   classification, 38
Propositional learning systems, 237

Prototypes, 230
Pruning, 61, 63, 67–69, 96, 107, 109, 194
Pruning
   backward, 61, 64
Pruning
   cost complexity, 69
Pruning
   forward, 61
Purity, 61, 62
Purity
   measure, 61
Purity measure, 59

Quadisc, 22, 121, 170, 173, 193, 263
Quadiscr, 225, 226
Quadratic discriminant, 12, 17, 22, 27, 214
Quadratic discriminants, 193
Quadratic functions of attributes, 22

Radial Basis Function, 85
Radial basis function, 93, 126, 223, 263
Radial Basis Function Network, 93
RAMnets, 103
RBF, 12, 85, 93, 223, 263
Recurrent networks, 88
Recursive partitioning, 9, 12, 16
Reduced nearest neighbour, 35
Reference class, 26
regression tree, 260
Regularisation, 23
Relational learning, 241
RETIS, 260
RG, 56
Risk assessment, 132
Rule-based methods, 10, 220
Rule-learning, 50

Satellite image dataset, 121, 143, 173
Scaling parameter, 32
Scatterplot smoother, 39
SDratio, 113, 170
Secific-to-general, 54
Secondary attribute, 123
Segmentation dataset, 145, 218
Selector, 56
Shuttle, 107
Shuttle dataset, 154, 218
Simulated digits data, 45
Skew, 115
Skewness, 28, 115, 170
SMART, 39, 216, 224, 225, 263
Smoothing parameter, 32
Smoothing parameters, 214
SNR, 119
Specific-to-general, 54, 57, 58, 79
Speed, 7
Splitting criteria, 61
Splitting criteria, 61
Splitting criterion, 62, 67, 70, 76
Splus, 26
Statistical approaches to classification, 2
Statistical measures, 112, 169
StatLog, 1, 4
StatLog
  collection of data, 53
StatLog
  objectives, 4
StatLog
  preprocessing, 124
Stepwise selection, 11
Stochastic gradient, 93
Storage, 223
Structured induction, 83
Subset selection, 199
Sum of squares, 18
Supervised learning, 1, 6, 8, 85
Supervised networks, 86
Supervised vector, 102
Supervisor, 8
Symbolic learning, 52
Symbolic ML, 52

Taxonomic, 58
Taxonomy, 54, 57, 58, 79
Technical dataset, 120, 161, 218
Tertiary attribute, 123
Test environment, 214
Test set, 8, 17, 108
Three-Mile Island, 7
Tiling algorithm, 96

Time, 223, 224
Time to learn, 7
Time to test, 7
Train-and-test, 108
Training
  optimisation, 94
Training set, 8, 17, 35, 108
Transformation, 121
Transformation of attributes, 25
Transformations of variables, 11
Tree-learning, 50
Trees-into-rules, 79
Tsetse dataset, 167, 218
Tuning of parameters, 109

UK credit dataset, 121
Uniform distribution, 32
Univariate kurtosis, 116
Univariate skewness, 116
Universal approximators, 88
Universal Computers, 88
Unsupervised learning, 1, 6, 85, 101
Upstart, 96
User’s guide to algorithms, 214

Vector Quantizers, 101
Vehicle, 170
Vehicle dataset, 138
Vertebrate, 53, 54, 57
Vertebrate species, 57
Voronoi tessellation, 101

XpertRule, 65, 80, 82

Yardstick methods, 210

Zero variance, 22