

List of Tables

Table 1.1 Contact Lens Data	6
Table 1.2 Weather Data	10
Table 1.3 Weather Data with Some Numeric Attributes	11
Table 1.4 Iris Data	14
Table 1.5 CPU Performance Data	16
Table 1.6 Labor Negotiations Data	17
Table 1.7 Soybean Data	20
Table 2.1 Iris Data as a Clustering Problem	41
Table 2.2 Weather Data with a Numeric Class	42
Table 2.3 Family Tree	44
Table 2.4 Sister-of Relation	45
Table 2.5 Another Relation	47
Table 3.1 New Iris Flower	73
Table 3.2 Training Data for the Shapes Problem	76
Table 4.1 Evaluating Attributes in the Weather Data	87
Table 4.2 Weather Data with Counts and Probabilities	91
Table 4.3 A New Day	92
Table 4.4 Numeric Weather Data with Summary Statistics	95
Table 4.5 Another New Day	96
Table 4.6 Weather Data with Identification Codes	106
Table 4.7 Gain Ratio Calculations for Figure 4.2 Tree Stumps	107
Table 4.8 Part of Contact Lens Data for which <i>astigmatism</i> = <i>yes</i>	112
Table 4.9 Part of Contact Lens Data for which <i>astigmatism</i> = <i>yes</i> and <i>tear production rate</i> = <i>normal</i>	113
Table 4.10 Item Sets for Weather Data with Coverage 2 or Greater	117
Table 4.11 Association Rules for Weather Data	120
Table 5.1 Confidence Limits for Normal Distribution	152
Table 5.2 Confidence Limits for Student's Distribution with 9 Degrees of Freedom	159
Table 5.3 Different Outcomes of a Two-Class Prediction	164
Table 5.4 Different Outcomes of a Three-Class Prediction	165
Table 5.5 Default Cost Matrixes	166
Table 5.6 Data for a Lift Chart	169
Table 5.7 Different Measures Used to Evaluate the False Positive versus False Negative Trade-Off	176
Table 5.8 Performance Measures for Numeric Prediction	180
Table 5.9 Performance Measures for Four Numeric Prediction Models	182
Table 6.1 Preparing Weather Data for Insertion into an <i>FP</i> -Tree	217
Table 6.2 Linear Models in the Model Tree	257
Table 7.1 First Five Instances from CPU Performance Data	327
Table 7.2 Transforming a Multiclass Problem into a Two-Class One	340

Table 7.3 Nested Dichotomy in the Form of a Code Matrix	342
Table 9.1 Top 10 Algorithms in Data Mining	376
Table 11.1 Unsupervised Attribute Filters	433
Table 11.2 Unsupervised Instance Filters	441
Table 11.3 Supervised Attribute Filters	444
Table 11.4 Supervised Instance Filters	444
Table 11.5 Classifier Algorithms in Weka	446
Table 11.6 Metalearning Algorithms in Weka	475
Table 11.7 Clustering Algorithms	480
Table 11.8 Association-Rule Learners	486
Table 11.9 Attribute Evaluation Methods for Attribute Selection	489
Table 11.10 Search Methods for Attribute Selection	490
Table 12.1 Visualization and Evaluation Components	499
Table 14.1 Generic Options for Learning Schemes	527
Table 14.2 Scheme-Specific Options for the J4.8 Decision Tree Learner	528
Table 16.1 Simple Learning Schemes in Weka	540
Table 17.1 Accuracy Obtained Using <i>IBk</i> , for Different Attribute Subsets	568
Table 17.2 Effect of Class Noise on <i>IBk</i> , for Different Neighborhood Sizes	569
Table 17.3 Effect of Training Set Size on <i>IBk</i> and <i>J48</i>	570
Table 17.4 Training Documents	580
Table 17.5 Test Documents	580
Table 17.6 Number of Rules for Different Values of Minimum Confidence and Support	584