

# List of Figures

Figure 1.1 Rules for the contact lens data.	12
Figure 1.2 Decision tree for the contact lens data.	13
Figure 1.3 Decision trees for the labor negotiations data.	18
Figure 2.1 A family tree and two ways of expressing the sister-of relation.	43
Figure 2.2 ARFF file for the weather data.	53
Figure 2.3 Multi-instance ARFF file for the weather data.	55
Figure 3.1 A linear regression function for the CPU performance data.	62
Figure 3.2 A linear decision boundary separating <i>Iris setosas</i> from <i>Iris versicolors</i> .	63
Figure 3.3 Constructing a decision tree interactively.	66
Figure 3.4 Models for the CPU performance data.	68
Figure 3.5 Decision tree for a simple disjunction.	69
Figure 3.6 The exclusive-or problem.	70
Figure 3.7 Decision tree with a replicated subtree.	71
Figure 3.8 Rules for the iris data.	74
Figure 3.9 The shapes problem.	76
Figure 3.10 Different ways of partitioning the instance space.	80
Figure 3.11 Different ways of representing clusters.	82
Figure 4.1 Pseudocode for 1R.	86
Figure 4.2 Tree stumps for the weather data.	100
Figure 4.3 Expanded tree stumps for the weather data.	102
Figure 4.4 Decision tree for the weather data.	103
Figure 4.5 Tree stump for the <i>ID code</i> attribute.	105
Figure 4.6 Covering algorithm.	109
Figure 4.7 The instance space during operation of a covering algorithm.	110
Figure 4.8 Pseudocode for a basic rule learner.	114
Figure 4.9 Logistic regression.	127
Figure 4.10 The perceptron.	129
Figure 4.11 The Winnow algorithm.	130
Figure 4.12 A <i>kD</i> -tree for four training instances.	133
Figure 4.13 Using a <i>kD</i> -tree to find the nearest neighbor of the star.	134
Figure 4.14 Ball tree for 16 training instances.	136
Figure 4.15 Ruling out an entire ball (gray) based on a target point (star) and its current nearest neighbor.	137
Figure 4.16 A ball tree.	141
Figure 5.1 A hypothetical lift chart.	170
Figure 5.2 Analyzing the expected benefit of a mailing campaign.	171
Figure 5.3 A sample ROC curve.	173
Figure 5.4 ROC curves for two learning schemes.	174
Figure 5.5 Effect of varying the probability threshold.	178
Figure 6.1 Example of subtree raising.	196

Figure 6.2 Pruning the labor negotiations decision tree.	200
Figure 6.3 Algorithm for forming rules by incremental reduced-error pruning.	207
Figure 6.4 RIPPER.	209
Figure 6.5 Algorithm for expanding examples into a partial tree.	210
Figure 6.6 Example of building a partial tree.	211
Figure 6.7 Rules with exceptions for the iris data.	213
Figure 6.8 Extended prefix trees for the weather data.	220
Figure 6.9 A maximum-margin hyperplane.	225
Figure 6.10 Support vector regression.	228
Figure 6.11 Example datasets and corresponding perceptrons.	233
Figure 6.12 Step versus sigmoid.	240
Figure 6.13 Gradient descent using the error function $w^2 + 1$ .	240
Figure 6.14 Multilayer perceptron with a hidden layer.	241
Figure 6.15 Hinge, squared, and 0 – 1 loss functions.	242
Figure 6.16 A boundary between two rectangular classes.	248
Figure 6.17 Pseudocode for model tree induction.	255
Figure 6.18 Model tree for a dataset with nominal attributes.	256
Figure 6.19 A simple Bayesian network for the weather data.	262
Figure 6.20 Another Bayesian network for the weather data.	264
Figure 6.21 The weather data.	270
Figure 6.22 Hierarchical clustering displays.	276
Figure 6.23 Clustering the weather data.	279
Figure 6.24 Hierarchical clusterings of the iris data.	281
Figure 6.25 A two-class mixture model.	285
Figure 6.26 <i>DensiTree</i> showing possible hierarchical clusterings of a given dataset.	291
Figure 7.1 Attribute space for the weather dataset.	311
Figure 7.2 Discretizing the <i>temperature</i> attribute using the entropy method.	318
Figure 7.3 The result of discretizing the <i>temperature</i> attribute.	318
Figure 7.4 Class distribution for a two-class, two-attribute problem.	321
Figure 7.5 Principal components transform of a dataset.	325
Figure 7.6 Number of international phone calls from Belgium, 1950–1973.	333
Figure 7.7 Overoptimistic probability estimation for a two-class problem.	344
Figure 8.1 Algorithm for bagging.	355
Figure 8.2 Algorithm for boosting.	359
Figure 8.3 Algorithm for additive logistic regression.	365
Figure 8.4 Simple option tree for the weather data.	366
Figure 8.5 Alternating decision tree for the weather data.	367
Figure 9.1 A tangled “web.”	391
Figure 11.1 The Explorer interface.	408
Figure 11.2 Weather data.	409
Figure 11.3 The Weka Explorer.	410

Figure 11.4 Using <i>J4.8</i> .	411
Figure 11.5 Output from the <i>J4.8</i> decision tree learner.	412
Figure 11.6 Visualizing the result of <i>J4.8</i> on the iris dataset.	415
Figure 11.7 Generic Object Editor.	417
Figure 11.8 The <i>SQLViewer</i> tool.	418
Figure 11.9 Choosing a filter.	420
Figure 11.10 The weather data with two attributes removed.	422
Figure 11.11 Processing the CPU performance data with <i>M5'</i> .	423
Figure 11.12 Output from the <i>M5'</i> program for numeric prediction.	425
Figure 11.13 Visualizing the errors.	426
Figure 11.14 Working on the segment-challenge data with the User Classifier.	428
Figure 11.15 Configuring a metalearner for boosting decision stumps.	429
Figure 11.16 Output from the <i>Apriori</i> program for association rules.	430
Figure 11.17 Visualizing the iris dataset.	431
Figure 11.18 Using Weka's metalearner for discretization.	443
Figure 11.19 Output of <i>NaiveBayes</i> on the weather data.	452
Figure 11.20 Visualizing a Bayesian network for the weather data (nominal version).	454
Figure 11.21 Changing the parameters for <i>J4.8</i> .	455
Figure 11.22 Output of <i>OneR</i> on the labor negotiations data.	458
Figure 11.23 Output of <i>PART</i> for the labor negotiations data.	460
Figure 11.24 Output of <i>SimpleLinearRegression</i> for the CPU performance data.	461
Figure 11.25 Output of <i>SMO</i> on the iris data.	463
Figure 11.26 Output of <i>SMO</i> with a nonlinear kernel on the iris data.	465
Figure 11.27 Output of <i>Logistic</i> on the iris data.	468
Figure 11.28 Using Weka's neural-network graphical user interface.	470
Figure 11.29 Output of <i>SimpleKMeans</i> on the weather data.	481
Figure 11.30 Output of <i>EM</i> on the weather data.	482
Figure 11.31 Clusters formed by <i>DBScan</i> on the iris data.	484
Figure 11.32 <i>OPTICS</i> visualization for the iris data.	485
Figure 11.33 Attribute selection: specifying an evaluator and a search method.	488
Figure 12.1 The Knowledge Flow interface.	496
Figure 12.2 Configuring a data source.	497
Figure 12.3 Status area after executing the configuration shown in Figure 12.1.	497
Figure 12.4 Operations on the Knowledge Flow components.	500
Figure 12.5 A Knowledge Flow that operates incrementally.	503
Figure 13.1 An experiment.	506
Figure 13.2 Statistical test results for the experiment in Figure 13.1.	509
Figure 13.3 Setting up an experiment in advanced mode.	511
Figure 13.4 An experiment in clustering.	513

Figure 13.5 Rows and columns of Figure 13.2.	514
Figure 14.1 Using Javadoc.	521
Figure 14.2 <i>DecisionStump</i> , a class of the <i>weka.classifiers.trees</i> package.	524
Figure 15.1 Source code for the message classifier.	532
Figure 16.1 Source code for the ID3 decision tree learner.	541
Figure 16.2 Source code produced by <i>weka.classifiers.trees.Id3</i> for the weather data.	551
Figure 16.3 Javadoc for the <i>Capability</i> enumeration.	556
Figure 17.1 The data viewer.	560
Figure 17.2 Output after building and testing the classifier.	564
Figure 17.3 The decision tree that has been built.	565