

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>A Simple Machine-Learning Task</b>                           | <b>1</b>  |
| 1.1      | Training Sets and Classifiers                                   | 1         |
| 1.2      | Minor Digression: Hill-Climbing Search                          | 5         |
| 1.3      | Hill Climbing in Machine Learning                               | 8         |
| 1.4      | The Induced Classifier's Performance                            | 11        |
| 1.5      | Some Difficulties with Available Data                           | 13        |
| 1.6      | Summary and Historical Remarks                                  | 15        |
| 1.7      | Solidify Your Knowledge   | 16        |
| <b>2</b> | <b>Probabilities: Bayesian Classifiers</b>                      | <b>19</b> |
| 2.1      | The Single-Attribute Case                                       | 19        |
| 2.2      | Vectors of Discrete Attributes                                  | 22        |
| 2.3      | Probabilities of Rare Events: Exploiting the Expert's Intuition | 26        |
| 2.4      | How to Handle Continuous Attributes                             | 30        |
| 2.5      | Gaussian "Bell" Function: A Standard <i>pdf</i>                 | 33        |
| 2.6      | Approximating PDFs with Sets of Gaussians                       | 34        |
| 2.7      | Summary and Historical Remarks                                  | 36        |
| 2.8      | Solidify Your Knowledge   | 40        |
| <b>3</b> | <b>Similarities: Nearest-Neighbor Classifiers</b>               | <b>43</b> |
| 3.1      | The $k$ -Nearest-Neighbor Rule                                  | 43        |
| 3.2      | Measuring Similarity  | 46        |
| 3.3      | Irrelevant Attributes and Scaling Problems                      | 49        |
| 3.4      | Performance Considerations                                      | 52        |
| 3.5      | Weighted Nearest Neighbors                                      | 55        |
| 3.6      | Removing Dangerous Examples                                     | 57        |
| 3.7      | Removing Redundant Examples                                     | 59        |
| 3.8      | Summary and Historical Remarks                                  | 61        |
| 3.9      | Solidify Your Knowledge   | 62        |

|          |  |     |
|----------|--|-----|
| <b>4</b> | <b>Inter-Class Boundaries: Linear and Polynomial Classifiers</b> | 65  |
| 4.1      | The Essence  | 65  |
| 4.2      | The Additive Rule: Perceptron Learning                           | 69  |
| 4.3      | The Multiplicative Rule: WINNOW                                  | 73  |
| 4.4      | Domains with More Than Two Classes                               | 76  |
| 4.5      | Polynomial Classifiers   | 79  |
| 4.6      | Specific Aspects of Polynomial Classifiers                       | 81  |
| 4.7      | Numerical Domains and Support Vector Machines                    | 84  |
| 4.8      | Summary and Historical Remarks                                   | 86  |
| 4.9      | Solidify Your Knowledge  | 87  |
| <b>5</b> | <b>Artificial Neural Networks</b>                                | 91  |
| 5.1      | Multilayer Perceptrons as Classifiers                            | 91  |
| 5.2      | Neural Network's Error   | 95  |
| 5.3      | Backpropagation of Error   | 97  |
| 5.4      | Special Aspects of Multilayer Perceptrons                        | 100 |
| 5.5      | Architectural Issues   | 104 |
| 5.6      | Radial-Basis Function Networks                                   | 106 |
| 5.7      | Summary and Historical Remarks                                   | 109 |
| 5.8      | Solidify Your Knowledge  | 110 |
| <b>6</b> | <b>Decision Trees</b>  | 113 |
| 6.1      | Decision Trees as Classifiers                                    | 113 |
| 6.2      | Induction of Decision Trees                                      | 117 |
| 6.3      | How Much Information Does an Attribute Convey?                   | 119 |
| 6.4      | Binary Split of a Numeric Attribute                              | 122 |
| 6.5      | Pruning  | 126 |
| 6.6      | Converting the Decision Tree into Rules                          | 130 |
| 6.7      | Summary and Historical Remarks                                   | 132 |
| 6.8      | Solidify Your Knowledge  | 133 |
| <b>7</b> | <b>Computational Learning Theory</b>                             | 137 |
| 7.1      | PAC Learning   | 137 |
| 7.2      | Examples of PAC Learnability                                     | 141 |
| 7.3      | Some Practical and Theoretical Consequences                      | 143 |
| 7.4      | VC-Dimension and Learnability                                    | 145 |
| 7.5      | Summary and Historical Remarks                                   | 148 |
| 7.6      | Exercises and Thought Experiments                                | 149 |
| <b>8</b> | <b>A Few Instructive Applications</b>                            | 151 |
| 8.1      | Character Recognition  | 151 |
| 8.2      | Oil-Spill Recognition  | 155 |
| 8.3      | Sleep Classification   | 158 |
| 8.4      | Brain-Computer Interface   | 161 |
| 8.5      | Medical Diagnosis  | 165 |

|           |   |            |
|-----------|---|------------|
| 8.6       | Text Classification .....                               | 167        |
| 8.7       | Summary and Historical Remarks .....                    | 169        |
| 8.8       | Exercises and Thought Experiments .....                 | 170        |
| <b>9</b>  | <b>Induction of Voting Assemblies .....</b>             | <b>173</b> |
| 9.1       | Bagging .....   | 173        |
| 9.2       | Schapire's Boosting .....                               | 176        |
| 9.3       | Adaboost: Practical Version of Boosting .....           | 179        |
| 9.4       | Variations on the Boosting Theme .....                  | 183        |
| 9.5       | Cost-Saving Benefits of the Approach .....              | 185        |
| 9.6       | Summary and Historical Remarks .....                    | 187        |
| 9.7       | Solidify Your Knowledge .....                           | 188        |
| <b>10</b> | <b>Some Practical Aspects to Know About .....</b>       | <b>191</b> |
| 10.1      | A Learner's Bias .....                                  | 191        |
| 10.2      | Imbalanced Training Sets .....                          | 194        |
| 10.3      | Context-Dependent Domains .....                         | 199        |
| 10.4      | Unknown Attribute Values .....                          | 202        |
| 10.5      | Attribute Selection .....                               | 204        |
| 10.6      | Miscellaneous .....                                     | 206        |
| 10.7      | Summary and Historical Remarks .....                    | 208        |
| 10.8      | Solidify Your Knowledge .....                           | 208        |
| <b>11</b> | <b>Performance Evaluation .....</b>                     | <b>211</b> |
| 11.1      | Basic Performance Criteria .....                        | 211        |
| 11.2      | Precision and Recall .....                              | 214        |
| 11.3      | Other Ways to Measure Performance .....                 | 219        |
| 11.4      | Learning Curves and Computational Costs .....           | 222        |
| 11.5      | Methodologies of Experimental Evaluation .....          | 224        |
| 11.6      | Summary and Historical Remarks .....                    | 227        |
| 11.7      | Solidify Your Knowledge .....                           | 228        |
| <b>12</b> | <b>Statistical Significance .....</b>                   | <b>231</b> |
| 12.1      | Sampling a Population .....                             | 231        |
| 12.2      | Benefiting from the Normal Distribution .....           | 235        |
| 12.3      | Confidence Intervals .....                              | 239        |
| 12.4      | Statistical Evaluation of a Classifier .....            | 241        |
| 12.5      | Another Kind of Statistical Evaluation .....            | 244        |
| 12.6      | Comparing Machine-Learning Techniques .....             | 245        |
| 12.7      | Summary and Historical Remarks .....                    | 247        |
| 12.8      | Solidify Your Knowledge .....                           | 248        |
| <b>13</b> | <b>Induction in Multi-Label Domains .....</b>           | <b>251</b> |
| 13.1      | Classical Machine Learning in Multi-Label Domains ..... | 251        |
| 13.2      | Treating Each Class Separately: Binary Relevance .....  | 254        |
| 13.3      | Classifier Chains .....                                 | 256        |

<http://www.springer.com/978-3-319-63912-3>

An Introduction to Machine Learning

Kubat, M.

2017, XIII, 348 p. 85 illus., 3 illus. in color., Hardcover

ISBN: 978-3-319-63912-3