

# Table of Contents

|                                                          |    |
|----------------------------------------------------------|----|
| <b>1. The Biological Paradigm</b>                        | 3  |
| 1.1 Neural computation                                   | 3  |
| 1.1.1 Natural and artificial neural networks             | 3  |
| 1.1.2 Models of computation                              | 5  |
| 1.1.3 Elements of a computing model                      | 9  |
| 1.2 Networks of neurons                                  | 9  |
| 1.2.1 Structure of the neurons                           | 9  |
| 1.2.2 Transmission of information                        | 11 |
| 1.2.3 Information processing at the neurons and synapses | 18 |
| 1.2.4 Storage of information – learning                  | 20 |
| 1.2.5 The neuron – a self-organizing system              | 21 |
| 1.3 Artificial neural networks                           | 23 |
| 1.3.1 Networks of primitive functions                    | 23 |
| 1.3.2 Approximation of functions                         | 24 |
| 1.3.3 Caveat                                             | 26 |
| 1.4 Historical and bibliographical remarks               | 26 |
| <b>2. Threshold Logic</b>                                | 29 |
| 2.1 Networks of functions                                | 29 |
| 2.1.1 Feed-forward and recurrent networks                | 29 |
| 2.1.2 The computing units                                | 31 |
| 2.2 Synthesis of Boolean functions                       | 33 |
| 2.2.1 Conjunction, disjunction, negation                 | 33 |
| 2.2.2 Geometric interpretation                           | 34 |
| 2.2.3 Constructive synthesis                             | 36 |
| 2.3 Equivalent networks                                  | 38 |
| 2.3.1 Weighted and unweighted networks                   | 39 |
| 2.3.2 Absolute and relative inhibition                   | 40 |
| 2.3.3 Binary signals and pulse coding                    | 41 |
| 2.4 Recurrent networks                                   | 42 |
| 2.4.1 Stored state networks                              | 42 |
| 2.4.2 Finite automata                                    | 43 |
| 2.4.3 Finite automata and recurrent networks             | 44 |
| 2.4.4 A first classification of neural networks          | 46 |

|           |                                                    |           |
|-----------|----------------------------------------------------|-----------|
| 2.5       | Harmonic analysis of logical functions .....       | 47        |
| 2.5.1     | General expression .....                           | 47        |
| 2.5.2     | The Hadamard-Walsh transform .....                 | 49        |
| 2.5.3     | Applications of threshold logic .....              | 50        |
| 2.6       | Historical and bibliographical remarks.....        | 52        |
| <b>3.</b> | <b>Weighted Networks – The Perceptron</b> .....    | <b>55</b> |
| 3.1       | Perceptrons and parallel processing .....          | 55        |
| 3.1.1     | Perceptrons as weighted threshold elements .....   | 55        |
| 3.1.2     | Computational limits of the perceptron model ..... | 57        |
| 3.2       | Implementation of logical functions .....          | 60        |
| 3.2.1     | Geometric interpretation .....                     | 60        |
| 3.2.2     | The XOR problem.....                               | 62        |
| 3.3       | Linearly separable functions .....                 | 63        |
| 3.3.1     | Linear separability.....                           | 63        |
| 3.3.2     | Duality of input space and weight space .....      | 64        |
| 3.3.3     | The error function in weight space .....           | 65        |
| 3.3.4     | General decision curves .....                      | 65        |
| 3.4       | Applications and biological analogy .....          | 66        |
| 3.4.1     | Edge detection with perceptrons .....              | 66        |
| 3.4.2     | The structure of the retina .....                  | 68        |
| 3.4.3     | Pyramidal networks and the neocognitron .....      | 69        |
| 3.4.4     | The silicon retina.....                            | 74        |
| 3.5       | Historical and bibliographical remarks.....        | 75        |
| <b>4.</b> | <b>Perceptron Learning</b> .....                   | <b>77</b> |
| 4.1       | Learning algorithms for neural networks .....      | 77        |
| 4.1.1     | Classes of learning algorithms .....               | 78        |
| 4.1.2     | Vector notation .....                              | 79        |
| 4.1.3     | Absolute linear separability .....                 | 80        |
| 4.1.4     | The error surface and the search method .....      | 81        |
| 4.2       | Algorithmic learning .....                         | 84        |
| 4.2.1     | Geometric visualization .....                      | 85        |
| 4.2.2     | Convergence of the algorithm .....                 | 87        |
| 4.2.3     | Accelerating convergence .....                     | 89        |
| 4.2.4     | The pocket algorithm .....                         | 90        |
| 4.2.5     | Complexity of perceptron learning .....            | 91        |
| 4.3       | Linear programming .....                           | 92        |
| 4.3.1     | Inner points of polytopes .....                    | 92        |
| 4.3.2     | Linear separability as linear optimization .....   | 94        |
| 4.3.3     | Karmarkar's algorithm .....                        | 95        |
| 4.4       | Historical and bibliographical remarks.....        | 97        |

|                                                           |     |
|-----------------------------------------------------------|-----|
| <b>5. Unsupervised Learning and Clustering Algorithms</b> | 99  |
| 5.1 Competitive learning                                  | 99  |
| 5.1.1 Generalization of the perceptron problem            | 99  |
| 5.1.2 Unsupervised learning through competition           | 101 |
| 5.2 Convergence analysis                                  | 103 |
| 5.2.1 The one-dimensional case – energy function          | 103 |
| 5.2.2 Multidimensional case – the classical methods       | 106 |
| 5.2.3 Unsupervised learning as minimization problem       | 108 |
| 5.2.4 Stability of the solutions                          | 110 |
| 5.3 Principal component analysis                          | 112 |
| 5.3.1 Unsupervised reinforcement learning                 | 112 |
| 5.3.2 Convergence of the learning algorithm               | 115 |
| 5.3.3 Multiple principal components                       | 117 |
| 5.4 Some applications                                     | 117 |
| 5.4.1 Pattern recognition                                 | 118 |
| 5.4.2 Image compression                                   | 118 |
| 5.5 Historical and bibliographical remarks                | 120 |
| <b>6. One and Two Layered Networks</b>                    | 123 |
| 6.1 Structure and geometric visualization                 | 123 |
| 6.1.1 Network architecture                                | 123 |
| 6.1.2 The XOR problem revisited                           | 124 |
| 6.1.3 Geometric visualization                             | 127 |
| 6.2 Counting regions in input and weight space            | 129 |
| 6.2.1 Weight space regions for the XOR problem            | 129 |
| 6.2.2 Bipolar vectors                                     | 131 |
| 6.2.3 Projection of the solution regions                  | 132 |
| 6.2.4 Geometric interpretation                            | 135 |
| 6.3 Regions for two layered networks                      | 138 |
| 6.3.1 Regions in weight space for the XOR problem         | 138 |
| 6.3.2 Number of regions in general                        | 139 |
| 6.3.3 Consequences                                        | 142 |
| 6.3.4 The Vapnik–Chervonenkis dimension                   | 142 |
| 6.3.5 The problem of local minima                         | 145 |
| 6.4 Historical and bibliographical remarks                | 147 |
| <b>7. The Backpropagation Algorithm</b>                   | 149 |
| 7.1 Learning as gradient descent                          | 149 |
| 7.1.1 Differentiable activation functions                 | 149 |
| 7.1.2 Regions in input space                              | 151 |
| 7.1.3 Local minima of the error function                  | 152 |
| 7.2 General feed-forward networks                         | 153 |
| 7.2.1 The learning problem                                | 153 |
| 7.2.2 Derivatives of network functions                    | 155 |
| 7.2.3 Steps of the backpropagation algorithm              | 159 |

|       |                                                        |     |
|-------|--------------------------------------------------------|-----|
| 7.2.4 | Learning with backpropagation .....                    | 161 |
| 7.3   | The case of layered networks .....                     | 162 |
| 7.3.1 | Extended network .....                                 | 162 |
| 7.3.2 | Steps of the algorithm .....                           | 164 |
| 7.3.3 | Backpropagation in matrix form .....                   | 168 |
| 7.3.4 | The locality of backpropagation .....                  | 169 |
| 7.3.5 | Error during training .....                            | 171 |
| 7.4   | Recurrent networks .....                               | 171 |
| 7.4.1 | Backpropagation through time .....                     | 172 |
| 7.4.2 | Hidden Markov Models .....                             | 175 |
| 7.4.3 | Variational problems .....                             | 178 |
| 7.5   | Historical and bibliographical remarks .....           | 180 |
| 8.    | <b>Fast Learning Algorithms</b> .....                  | 183 |
| 8.1   | Introduction – classical backpropagation .....         | 183 |
| 8.1.1 | Backpropagation with momentum .....                    | 184 |
| 8.1.2 | The fractal geometry of backpropagation .....          | 190 |
| 8.2   | Some simple improvements to backpropagation .....      | 197 |
| 8.2.1 | Initial weight selection .....                         | 197 |
| 8.2.2 | Clipped derivatives and offset term .....              | 199 |
| 8.2.3 | Reducing the number of floating-point operations ..... | 200 |
| 8.2.4 | Data decorrelation .....                               | 202 |
| 8.3   | Adaptive step algorithms .....                         | 204 |
| 8.3.1 | Silva and Almeida's algorithm .....                    | 205 |
| 8.3.2 | Delta-bar-delta .....                                  | 207 |
| 8.3.3 | Rprop .....                                            | 208 |
| 8.3.4 | The Dynamic Adaption algorithm .....                   | 209 |
| 8.4   | Second-order algorithms .....                          | 210 |
| 8.4.1 | Quickprop .....                                        | 211 |
| 8.4.2 | QRprop .....                                           | 212 |
| 8.4.3 | Second-order backpropagation .....                     | 214 |
| 8.5   | Relaxation methods .....                               | 221 |
| 8.5.1 | Weight and node perturbation .....                     | 221 |
| 8.5.2 | Symmetric and asymmetric relaxation .....              | 222 |
| 8.5.3 | A final thought on taxonomy .....                      | 223 |
| 8.6   | Historical and bibliographical remarks .....           | 224 |
| 9.    | <b>Statistics and Neural Networks</b> .....            | 227 |
| 9.1   | Linear and nonlinear regression .....                  | 227 |
| 9.1.1 | The problem of good generalization .....               | 227 |
| 9.1.2 | Linear regression .....                                | 229 |
| 9.1.3 | Nonlinear units .....                                  | 231 |
| 9.1.4 | Computing the prediction error .....                   | 233 |
| 9.1.5 | The jackknife and cross-validation .....               | 236 |
| 9.1.6 | Committees of networks .....                           | 237 |

|            |                                                       |            |
|------------|-------------------------------------------------------|------------|
| 9.2        | Multiple regression .....                             | 240        |
| 9.2.1      | Visualization of the solution regions .....           | 240        |
| 9.2.2      | Linear equations and the pseudoinverse .....          | 242        |
| 9.2.3      | The hidden layer .....                                | 243        |
| 9.2.4      | Computation of the pseudoinverse .....                | 244        |
| 9.3        | Classification networks .....                         | 245        |
| 9.3.1      | An application: NETtalk .....                         | 246        |
| 9.3.2      | The Bayes property of classifier networks .....       | 247        |
| 9.3.3      | Connectionist speech recognition .....                | 250        |
| 9.3.4      | Autoregressive models for time series analysis .....  | 258        |
| 9.4        | Historical and bibliographical remarks .....          | 259        |
| <b>10.</b> | <b>The Complexity of Learning .....</b>               | <b>263</b> |
| 10.1       | Network functions .....                               | 263        |
| 10.1.1     | Learning algorithms for multilayer networks .....     | 263        |
| 10.1.2     | Hilbert's problem and computability .....             | 263        |
| 10.1.3     | Kolmogorov's theorem .....                            | 265        |
| 10.2       | Function approximation .....                          | 267        |
| 10.2.1     | The one-dimensional case .....                        | 267        |
| 10.2.2     | The multidimensional case .....                       | 269        |
| 10.3       | Complexity of learning problems .....                 | 271        |
| 10.3.1     | Complexity classes .....                              | 272        |
| 10.3.2     | <i>NP</i> -complete learning problems .....           | 275        |
| 10.3.3     | Complexity of learning with AND-OR networks .....     | 277        |
| 10.3.4     | Simplifications of the network architecture .....     | 280        |
| 10.3.5     | Learning with hints .....                             | 281        |
| 10.4       | Historical and bibliographical remarks .....          | 284        |
| <b>11.</b> | <b>Fuzzy Logic .....</b>                              | <b>287</b> |
| 11.1       | Fuzzy sets and fuzzy logic .....                      | 287        |
| 11.1.1     | Imprecise data and imprecise rules .....              | 287        |
| 11.1.2     | The fuzzy set concept .....                           | 288        |
| 11.1.3     | Geometric representation of fuzzy sets .....          | 290        |
| 11.1.4     | Fuzzy set theory, logic operators, and geometry ..... | 294        |
| 11.1.5     | Families of fuzzy operators .....                     | 295        |
| 11.2       | Fuzzy inferences .....                                | 299        |
| 11.2.1     | Inferences from imprecise data .....                  | 299        |
| 11.2.2     | Fuzzy numbers and inverse operation .....             | 300        |
| 11.3       | Control with fuzzy logic .....                        | 302        |
| 11.3.1     | Fuzzy controllers .....                               | 302        |
| 11.3.2     | Fuzzy networks .....                                  | 303        |
| 11.3.3     | Function approximation with fuzzy methods .....       | 305        |
| 11.3.4     | The eye as a fuzzy system – color vision .....        | 306        |
| 11.4       | Historical and bibliographical remarks .....          | 307        |

|                                                             |     |
|-------------------------------------------------------------|-----|
| <b>12. Associative Networks</b>                             | 309 |
| 12.1 Associative pattern recognition                        | 309 |
| 12.1.1 Recurrent networks and types of associative memories | 309 |
| 12.1.2 Structure of an associative memory                   | 311 |
| 12.1.3 The eigenvector automaton                            | 312 |
| 12.2 Associative learning                                   | 314 |
| 12.2.1 Hebbian learning – the correlation matrix            | 314 |
| 12.2.2 Geometric interpretation of Hebbian learning         | 317 |
| 12.2.3 Networks as dynamical systems – some experiments     | 318 |
| 12.2.4 Another visualization                                | 322 |
| 12.3 The capacity problem                                   | 323 |
| 12.4 The pseudoinverse                                      | 324 |
| 12.4.1 Definition and properties of the pseudoinverse       | 325 |
| 12.4.2 Orthogonal projections                               | 327 |
| 12.4.3 Holographic memories                                 | 330 |
| 12.4.4 Translation invariant pattern recognition            | 331 |
| 12.5 Historical and bibliographical remarks                 | 333 |
| <b>13. The Hopfield Model</b>                               | 335 |
| 13.1 Synchronous and asynchronous networks                  | 335 |
| 13.1.1 Recursive networks with stochastic dynamics          | 335 |
| 13.1.2 The bidirectional associative memory                 | 336 |
| 13.1.3 The energy function                                  | 338 |
| 13.2 Definition of Hopfield networks                        | 339 |
| 13.2.1 Asynchronous networks                                | 339 |
| 13.2.2 Examples of the model                                | 341 |
| 13.2.3 Isomorphism between the Hopfield and Ising models    | 346 |
| 13.3 Converge to stable states                              | 347 |
| 13.3.1 Dynamics of Hopfield networks                        | 347 |
| 13.3.2 Convergence proof                                    | 348 |
| 13.3.3 Hebbian learning                                     | 352 |
| 13.4 Equivalence of Hopfield and perceptron learning        | 354 |
| 13.4.1 Perceptron learning in Hopfield networks             | 354 |
| 13.4.2 Complexity of learning in Hopfield models            | 356 |
| 13.5 Parallel combinatorics                                 | 357 |
| 13.5.1 <i>NP</i> -complete problems and massive parallelism | 357 |
| 13.5.2 The multiflop problem                                | 357 |
| 13.5.3 The eight rooks problem                              | 359 |
| 13.5.4 The eight queens problem                             | 360 |
| 13.5.5 The traveling salesman                               | 361 |
| 13.5.6 The limits of Hopfield networks                      | 363 |
| 13.6 Implementation of Hopfield networks                    | 365 |
| 13.6.1 Electrical implementation                            | 365 |
| 13.6.2 Optical implementation                               | 366 |
| 13.7 Historical and bibliographical remarks                 | 368 |

|                                                       |     |
|-------------------------------------------------------|-----|
| <b>14. Stochastic Networks</b>                        | 371 |
| 14.1 Variations of the Hopfield model                 | 371 |
| 14.1.1 The continuous model                           | 372 |
| 14.2 Stochastic systems                               | 373 |
| 14.2.1 Simulated annealing                            | 374 |
| 14.2.2 Stochastic neural networks                     | 375 |
| 14.2.3 Markov chains                                  | 376 |
| 14.2.4 The Boltzmann distribution                     | 379 |
| 14.2.5 Physical meaning of the Boltzmann distribution | 382 |
| 14.3 Learning algorithms and applications             | 383 |
| 14.3.1 Boltzmann learning                             | 383 |
| 14.3.2 Combinatorial optimization                     | 385 |
| 14.4 Historical and bibliographical remarks           | 386 |
| <b>15. Kohonen Networks</b>                           | 389 |
| 15.1 Self-organization                                | 389 |
| 15.1.1 Charting input space                           | 389 |
| 15.1.2 Topology preserving maps in the brain          | 390 |
| 15.2 Kohonen's model                                  | 393 |
| 15.2.1 Learning algorithm                             | 393 |
| 15.2.2 Mapping high-dimensional spaces                | 397 |
| 15.3 Analysis of convergence                          | 399 |
| 15.3.1 Potential function – the one-dimensional case  | 399 |
| 15.3.2 The two-dimensional case                       | 401 |
| 15.3.3 Effect of a unit's neighborhood                | 402 |
| 15.3.4 Metastable states                              | 403 |
| 15.3.5 What dimension for Kohonen networks?           | 405 |
| 15.4 Applications                                     | 406 |
| 15.4.1 Approximation of functions                     | 406 |
| 15.4.2 Inverse kinematics                             | 407 |
| 15.5 Historical and bibliographical remarks           | 409 |
| <b>16. Modular Neural Networks</b>                    | 411 |
| 16.1 Constructive algorithms for modular networks     | 411 |
| 16.1.1 Cascade correlation                            | 412 |
| 16.1.2 Optimal modules and mixtures of experts        | 413 |
| 16.2 Hybrid networks                                  | 414 |
| 16.2.1 The ART architectures                          | 414 |
| 16.2.2 Maximum entropy                                | 417 |
| 16.2.3 Counterpropagation networks                    | 418 |
| 16.2.4 Spline networks                                | 420 |
| 16.2.5 Radial basis functions                         | 422 |
| 16.3 Historical and bibliographical remarks           | 424 |

|                                                        |     |
|--------------------------------------------------------|-----|
| <b>17. Genetic Algorithms</b>                          | 427 |
| 17.1 Coding and operators                              | 427 |
| 17.1.1 Optimization problems                           | 427 |
| 17.1.2 Methods of stochastic optimization              | 428 |
| 17.1.3 Genetic coding                                  | 431 |
| 17.1.4 Information exchange with genetic operators     | 432 |
| 17.2 Properties of genetic algorithms                  | 433 |
| 17.2.1 Convergence analysis                            | 433 |
| 17.2.2 Deceptive problems                              | 437 |
| 17.2.3 Genetic drift                                   | 438 |
| 17.2.4 Gradient methods versus genetic algorithms      | 440 |
| 17.3 Neural networks and genetic algorithms            | 441 |
| 17.3.1 The problem of symmetries                       | 441 |
| 17.3.2 A numerical experiment                          | 443 |
| 17.3.3 Other applications of GAs                       | 444 |
| 17.4 Historical and bibliographical remarks            | 446 |
| <b>18. Hardware for Neural Networks</b>                | 449 |
| 18.1 Taxonomy of neural hardware                       | 449 |
| 18.1.1 Performance requirements                        | 449 |
| 18.1.2 Types of neurocomputers                         | 451 |
| 18.2 Analog neural networks                            | 452 |
| 18.2.1 Coding                                          | 453 |
| 18.2.2 VLSI transistor circuits                        | 454 |
| 18.2.3 Transistors with stored charge                  | 456 |
| 18.2.4 CCD components                                  | 457 |
| 18.3 Digital networks                                  | 459 |
| 18.3.1 Numerical representation of weights and signals | 459 |
| 18.3.2 Vector and signal processors                    | 460 |
| 18.3.3 Systolic arrays                                 | 461 |
| 18.3.4 One-dimensional structures                      | 463 |
| 18.4 Innovative computer architectures                 | 466 |
| 18.4.1 VLSI microprocessors for neural networks        | 466 |
| 18.4.2 Optical computers                               | 469 |
| 18.4.3 Pulse coded networks                            | 472 |
| 18.5 Historical and bibliographical remarks            | 474 |