

| | |
|--|----|
| Preface | xi |
| Contributing Authors | xv |
| 1 | |
| Neural Nets and Related Model Structures for Nonlinear System Identification | 1 |
| <i>Jonas Sjöberg and Lester S.H. Ngia</i> | |
| 1.1 Introduction | 2 |
| 1.2 Black-box Identification Approach | 3 |
| 1.3 Fitting the Parameters | 4 |
| 1.4 The Nonlinear Mapping | 5 |
| 1.5 Some Standard Theory on Model Properties | 7 |
| 1.5.1 Approximation Properties | 7 |
| 1.5.2 Bias-Variance | 8 |
| 1.6 Bias-Variance Trade-off | 9 |
| 1.6.1 Regularization | 10 |
| 1.6.2 Implicit Regularization, Stopped Iterations, and Overtraining | 11 |
| 1.6.3 Pruning and Shrinking | 12 |
| 1.7 Discussion | 12 |
| 1.8 Dynamic Systems | 13 |
| 1.9 A stepwise Approach to Nonlinear Black-box Identification | 15 |
| 1.10 Examples | 17 |
| 1.10.1 Nonlinear Modeling of Hybrid Echo in a Telephone Network | 17 |
| 1.10.2 Example of Physical Modeling: Aerodynamic Identification | 22 |
| 1.11 Conclusions | 25 |
| References | 26 |
| 2 | |
| Enhanced Multi-Stream Kalman Filter Training for Recurrent Networks | 29 |
| <i>Lee A. Feldkamp, Danil V. Prokhorov, Charles F. Eagen, and Fumin Yuan</i> | |
| 2.1 Introduction | 29 |
| 2.2 Network Architecture and Execution | 31 |
| 2.3 Gradient Calculation | 33 |
| 2.3.1 Traditional BPTT | 34 |
| | v |

| | | |
|--|---|----|
| 2.3.2 | Aggregate BPTT | 36 |
| 2.4 | EKF Multi-Stream Training | 37 |
| 2.4.1 | The Kalman Recursion | 37 |
| 2.4.2 | Multi-Stream Training | 38 |
| 2.5 | A Modeling Example | 40 |
| 2.6 | Alternating training of weights and initial states | 42 |
| 2.7 | Toward more coordinated training of weights and states—A Motivating Example | 44 |
| 2.8 | The Mackey-Glass Series | 46 |
| 2.9 | Use of a structured evolver | 49 |
| 2.10 | Summary | 51 |
| References | | 52 |
| 3 | | |
| The Support Vector Method of Function Estimation | | 55 |
| <i>Vladimir Vapnik</i> | | |
| 3.1 | Introduction | 55 |
| 3.2 | The Optimal Hyperplane | 56 |
| 3.2.1 | The Optimal Hyperplane for Separable Data | 56 |
| 3.2.2 | The Optimal Hyperplane for Nonseparable Data | 62 |
| 3.2.3 | Statistical Properties of the Optimal Hyperplane | 64 |
| 3.3 | The Support Vector Machine for Pattern Recognition | 66 |
| 3.3.1 | The Idea of Support Vector Machines | 66 |
| 3.3.2 | Generalization in High-Dimensional Space | 67 |
| 3.3.3 | Hilbert-Schmidt theory and Mercer's theorem | 67 |
| 3.3.4 | Constructing SV Machines | 68 |
| 3.3.5 | Selection of a SV Machine Using Bounds | 69 |
| 3.4 | Examples of SV Machines for Pattern Recognition | 70 |
| 3.4.1 | Polynomial Support Vector Machines | 71 |
| 3.4.2 | Radial Basis Function SV Machines | 71 |
| 3.4.3 | Two-Layer Neural SV Machines | 72 |
| 3.5 | The SVM for regression estimation | 73 |
| 3.5.1 | ϵ -Insensitive Loss-Functions | 73 |
| 3.5.2 | Minimizing the Risk with ϵ -insensitive Loss-function | 74 |
| 3.5.3 | SV Machines for Regression Estimation | 78 |
| 3.6 | Kernels for Estimating Real-Valued Functions | 78 |
| 3.6.1 | Kernels Generating Splines | 79 |
| 3.6.2 | Kernels Generating Splines with an Infinite Number of Knots | 80 |
| 3.6.3 | Kernels Generating Fourier Expansions | 81 |
| 3.7 | Solving Operator Equations | 82 |
| 3.8 | Conclusion | 84 |
| References | | 84 |
| 4 | | |
| Parametric Density Estimation for the Classification of Acoustic Feature Vectors in Speech Recognition | | 87 |
| <i>Sankar Basu and Charles A. Micchelli</i> | | |
| 4.1 | Introduction | 87 |
| 4.2 | The Probability Densities | 89 |
| 4.3 | Maximum Likelihood Estimation | 96 |

| | | |
|---|--|-----|
| 4.4 | Mixture Models | 105 |
| 4.5 | Numerical experiment | 112 |
| 4.6 | Conclusion | 114 |
| References | | 114 |
| 5 | | |
| Wavelet Based Modeling of Nonlinear Systems | | 119 |
| <i>Yi Yu, Wayne Lawton[†], Seng Luan Lee[†], Shaohua Tan[†], Joos Vandewalle[#]</i> | | |
| 5.1 | Introduction | 120 |
| 5.2 | Wavelet Interpolation Method | 123 |
| 5.2.1 | Wavelet-Based Norms | 123 |
| 5.2.2 | Interpolation Method | 126 |
| 5.2.3 | Interpolation Algorithm | 127 |
| 5.3 | Wavelet Interpolation in Sobolev Space | 129 |
| 5.3.1 | Wavelet-based Sobolev Norms | 129 |
| 5.3.2 | Interpolation Method | 130 |
| 5.3.3 | Spatial Adaptivity | 133 |
| 5.3.4 | Multi-scale Interpolation | 134 |
| 5.4 | Numerical Examples | 136 |
| 5.4.1 | Nonlinear System Modeling from Random Observations | 138 |
| 5.4.2 | Dynamical System Modeling from Time Series | 140 |
| 5.5 | Conclusions | 141 |
| Appendix: Constrained Minimal Norm Interpolation | | 143 |
| References | | 146 |
| 6 | | |
| Nonlinear Identification based on Fuzzy Models | | 149 |
| <i>Vincent Wertz* and Stephen Yurkovich**</i> | | |
| 6.1 | Introduction | 149 |
| 6.1.1 | Motivations | 149 |
| 6.1.2 | Chapter outline | 151 |
| 6.2 | Functional fuzzy models | 151 |
| 6.2.1 | Fuzzy models | 151 |
| 6.2.2 | Takagi-Sugeno fuzzy models | 154 |
| 6.2.3 | Relation with LTV-LPV models and control issues | 155 |
| 6.2.4 | Structure, nonlinear and linear parameters | 155 |
| 6.3 | Fuzzy Modelling algorithms without structure determination | 157 |
| 6.3.1 | Fixed shape for the membership functions | 157 |
| 6.3.2 | Estimation of nonlinear parameters - ANFIS | 158 |
| 6.4 | Structure determination | 160 |
| 6.4.1 | Clustering methods | 160 |
| 6.4.2 | Subtractive clustering | 163 |
| 6.5 | A worked example | 164 |
| 6.5.1 | The glass furnace process and data | 164 |
| 6.5.2 | Linear identification results | 166 |
| 6.5.3 | TS Fuzzy models | 167 |
| 6.6 | Discussion | 172 |
| References | | 174 |

| | | |
|--------|---|-----|
| 7 | Statistical Learning in Control and Matrix Theory | 177 |
| | <i>M. Vidyasagar</i> | |
| 7.1 | Introduction | 178 |
| 7.2 | Paradigm of Controller Synthesis Problem | 181 |
| 7.3 | Various Types of "Near" Minima | 184 |
| 7.4 | A General Approach to Randomized Algorithms | 187 |
| 7.4.1 | The UCEM Property | 187 |
| 7.4.2 | An Approach to Finding Approximate Near Minima with High Confidence | 188 |
| 7.4.3 | A Universal Algorithm for Finding Probable Near Minima | 190 |
| 7.4.4 | An Algorithm for Finding Probably Approximate Near Minima | 191 |
| 7.5 | Some Sufficient Conditions for the UCEM Property | 194 |
| 7.5.1 | Definitions of the VC-dimension and P-Dimension | 194 |
| 7.5.2 | Finiteness of the VC- and P-Dimensions Implies the UCEM Property | 195 |
| 7.5.3 | Upper Bounds for the VC-Dimension | 196 |
| 7.6 | Robust Stabilization | 198 |
| 7.7 | Weighted H_∞ -Norm Minimization | 201 |
| 7.8 | Randomized Algorithms for Matrix Problems | 201 |
| 7.9 | Sample Complexity Issues | 204 |
| 7.10 | Conclusions | 205 |
| | References | 206 |
| 8 | Nonlinear Time-Series Analysis | 209 |
| | <i>Ulrich Parlitz</i> | |
| 8.1 | Introduction | 209 |
| 8.2 | State space reconstruction | 210 |
| 8.3 | Influence of the reconstruction parameters | 213 |
| 8.3.1 | Choice of the delay time | 213 |
| 8.3.2 | Choice of the embedding dimension | 215 |
| 8.3.3 | Estimating reconstruction parameters | 215 |
| 8.4 | Noise reduction | 216 |
| 8.4.1 | Linear Filters | 216 |
| 8.4.2 | Nonlinear signal separation | 217 |
| 8.5 | Detecting nonlinearities | 217 |
| 8.6 | Modelling and prediction | 218 |
| 8.7 | Fractal dimensions | 219 |
| 8.7.1 | Correlation dimension | 219 |
| 8.7.2 | Information dimension | 219 |
| 8.8 | Lyapunov exponents | 221 |
| 8.8.1 | Jacobian-based methods | 222 |
| 8.8.2 | Direct methods | 222 |
| 8.9 | Synchronization of chaotic dynamics | 224 |
| 8.9.1 | Estimating model parameters | 224 |
| 8.9.2 | Generalized synchronization | 224 |
| 8.10 | Spatio-temporal time series | 225 |
| 8.10.1 | Linear decomposition into spatial modes | 226 |
| 8.10.2 | Local models | 227 |
| 8.10.3 | Numerical Example | 229 |

| | | |
|------------|---|-----|
| References | 232 | |
| 9 | The K.U.Leuven Time Series Prediction Competition | 241 |
| | <i>Johan A.K. Suykens and Joos Vandewalle</i> | |
| 9.1 | Introduction | 241 |
| 9.2 | Origin of the data | 242 |
| 9.3 | Results of the competition | 246 |
| 9.4 | Conclusions | 251 |
| References | 251 | |
| Index | 254 | |