Contents

| 16.2 | Regularization | 266 |
|-------------|--|-----|
| 16.3 | Pruning Methods | 268 |
| 16.4 | Constructive Methods | 268 |
| 16.5 | Weight Decay | 269 |
| 16.6 | Information Minimization | 271 |
| 16.7 | Replicated Networks | 272 |
| 16.8 | Training with Noisy Data | 273 |
| 16.9 | Use of Domain-Dependent Prior Information | 274 |
| 16.10 | Hint Functions | 275 |
| 16.11 | Knowledge-Based Neural Nets | 275 |
| 16.12 | Physical Models to Generate Additional Data | 276 |
| 17 | Effects of Training with Noisy Inputs | 277 |
| 17.1 | Convolution Property of Training with Jitter | 277 |
| 17.2 | Error Regularization and Training with Jitter | 281 |
| 17.3 | Training with Jitter and Sigmoid Scaling | 283 |
| 17.4 | Extension to General Layered Neural Networks | 288 |
| 17.5 | Remarks | 289 |
| 17.6 | Further Examples | 290 |
| A | Linear Regression | 293 |
| A.1 | Newton's Method | 294 |
| A.2 | Gradient Descent | 295 |
| A.3 | The LMS Algorithm | 298 |
| B | Principal Components Analysis | 299 |
| B .1 | Autoencoder Networks and Principal Components | 303 |
| B.2 | Discriminant Analysis Projections | 306 |
| С | Jitter Calculations | 311 |
| C.1 | Jitter: Small–Perturbation Approximation | 311 |
| C.2 | Jitter: CDF–PDF Convolution in <i>n</i> Dimensions | 311 |
| C.3 | Jitter: CDF-PDF Convolution in One Dimension | 314 |
| D | Sigmoid-like Nonlinear Functions | 315 |
| | References | 319 |
| | Index | 339 |

viii