2.3 Adaptive Learning: Concepts and Inductive Principles, 40
   2.3.1 Philosophy, Major Concepts, and Issues, 40
   2.3.2 A Priori Knowledge and Model Complexity, 43
   2.3.3 Inductive Principles, 45
   2.3.4 Alternative Learning Formulations, 55
2.4 Summary, 58

3 Regularization Framework 61
   3.1 Curse and Complexity of Dimensionality, 62
   3.2 Function Approximation and Characterization of Complexity, 66
   3.3 Penalization, 70
      3.3.1 Parametric Penalties, 72
      3.3.2 Nonparametric Penalties, 73
   3.4 Model Selection (Complexity Control), 73
      3.4.1 Analytical Model Selection Criteria, 75
      3.4.2 Model Selection via Resampling, 78
      3.4.3 Bias–Variance Tradeoff, 80
      3.4.4 Example of Model Selection, 85
      3.4.5 Function Approximation versus Predictive Learning, 88
   3.5 Summary, 96

4 Statistical Learning Theory 99
   4.1 Conditions for Consistency and Convergence of ERM, 101
   4.2 Growth Function and VC Dimension, 107
      4.2.1 VC Dimension for Classification and Regression Problems, 110
      4.2.2 Examples of Calculating VC Dimension, 111
   4.3 Bounds on the Generalization, 115
      4.3.1 Classification, 116
      4.3.2 Regression, 118
      4.3.3 Generalization Bounds and Sampling Theorem, 120
   4.4 Structural Risk Minimization, 122
      4.4.1 Dictionary Representation, 124
      4.4.2 Feature Selection, 125
      4.4.3 Penalization Formulation, 126
      4.4.4 Input Preprocessing, 126
      4.4.5 Initial Conditions for Training Algorithm, 127
   4.5 Comparisons of Model Selection for Regression, 128
      4.5.1 Model Selection for Linear Estimators, 134
      4.5.2 Model Selection for k-Nearest-Neighbor Regression, 137
      4.5.3 Model Selection for Linear Subset Regression, 140
      4.5.4 Discussion, 141
   4.6 Measuring the VC Dimension, 143
   4.7 VC Dimension, Occam’s Razor, and Popper’s Falsifiability, 146
   4.8 Summary and Discussion, 149
### 5 Nonlinear Optimization Strategies

5.1 Stochastic Approximation Methods, 154  
5.1.1 Linear Parameter Estimation, 155  
5.1.2 Backpropagation Training of MLP Networks, 156

5.2 Iterative Methods, 161  
5.2.1 EM Methods for Density Estimation, 161  
5.2.2 Generalized Inverse Training of MLP Networks, 164

5.3 Greedy Optimization, 169  
5.3.1 Neural Network Construction Algorithms, 169  
5.3.2 Classification and Regression Trees, 170

5.4 Feature Selection, Optimization, and Statistical Learning Theory, 173

5.5 Summary, 175

### 6 Methods for Data Reduction and Dimensionality Reduction

6.1 Vector Quantization and Clustering, 183  
6.1.1 Optimal Source Coding in Vector Quantization, 184  
6.1.2 Generalized Lloyd Algorithm, 187  
6.1.3 Clustering, 191  
6.1.4 EM Algorithm for VQ and Clustering, 192  
6.1.5 Fuzzy Clustering, 195

6.2 Dimensionality Reduction: Statistical Methods, 201  
6.2.1 Linear Principal Components, 202  
6.2.2 Principal Curves and Surfaces, 205  
6.2.3 Multidimensional Scaling, 209

6.3 Dimensionality Reduction: Neural Network Methods, 214  
6.3.1 Discrete Principal Curves and Self-Organizing Map Algorithm, 215  
6.3.2 Statistical Interpretation of the SOM Method, 218  
6.3.3 Flow-Through Version of the SOM and Learning Rate Schedules, 222  
6.3.4 SOM Applications and Modifications, 224  
6.3.5 Self-Supervised MLP, 230

6.4 Methods for Multivariate Data Analysis, 232  
6.4.1 Factor Analysis, 233  
6.4.2 Independent Component Analysis, 242

6.5 Summary, 247

### 7 Methods for Regression

7.1 Taxonomy: Dictionary versus Kernel Representation, 252

7.2 Linear Estimators, 256  
7.2.1 Estimation of Linear Models and Equivalence of Representations, 258  
7.2.2 Analytic Form of Cross-Validation, 262
7.2.3 Estimating Complexity of Penalized Linear Models, 263
7.2.4 Nonadaptive Methods, 269
7.3 Adaptive Dictionary Methods, 277
  7.3.1 Additive Methods and Projection Pursuit Regression, 279
  7.3.2 Multilayer Perceptrons and Backpropagation, 284
  7.3.3 Multivariate Adaptive Regression Splines, 293
  7.3.4 Orthogonal Basis Functions and Wavelet Signal Denoising, 298
7.4 Adaptive Kernel Methods and Local Risk Minimization, 309
  7.4.1 Generalized Memory-Based Learning, 313
  7.4.2 Constrained Topological Mapping, 314
7.5 Empirical Studies, 319
  7.5.1 Predicting Net Asset Value (NAV) of Mutual Funds, 320
  7.5.2 Comparison of Adaptive Methods for Regression, 326
7.6 Combining Predictive Models, 332
7.7 Summary, 337

8 Classification 340
  8.1 Statistical Learning Theory Formulation, 343
  8.2 Classical Formulation, 348
    8.2.1 Statistical Decision Theory, 348
    8.2.2 Fisher's Linear Discriminant Analysis, 362
  8.3 Methods for Classification, 366
    8.3.1 Regression-Based Methods, 368
    8.3.2 Tree-Based Methods, 378
    8.3.3 Nearest-Neighbor and Prototype Methods, 382
    8.3.4 Empirical Comparisons, 385
  8.4 Combining Methods and Boosting, 390
    8.4.1 Boosting as an Additive Model, 395
    8.4.2 Boosting for Regression Problems, 400
  8.5 Summary, 401

9 Support Vector Machines 404
  9.1 Motivation for Margin-Based Loss, 408
  9.2 Margin-Based Loss, Robustness, and Complexity Control, 414
  9.3 Optimal Separating Hyperplane, 418
  9.4 High-Dimensional Mapping and Inner Product Kernels, 426
  9.5 Support Vector Machine for Classification, 430
  9.6 Support Vector Implementations, 438
  9.7 Support Vector Regression, 439
  9.8 SVM Model Selection, 445
  9.9 Support Vector Machines and Regularization, 453