

6.3	Hidden Units	191
6.4	Architecture Design	197
6.5	Back-Propagation and Other Differentiation Algorithms	204
6.6	Historical Notes	224
7	Regularization for Deep Learning	228
7.1	Parameter Norm Penalties	230
7.2	Norm Penalties as Constrained Optimization	237
7.3	Regularization and Under-Constrained Problems	239
7.4	Dataset Augmentation	240
7.5	Noise Robustness	242
7.6	Semi-Supervised Learning	243
7.7	Multi-Task Learning	244
7.8	Early Stopping	246
7.9	Parameter Tying and Parameter Sharing	253
7.10	Sparse Representations	254
7.11	Bagging and Other Ensemble Methods	256
7.12	Dropout	258
7.13	Adversarial Training	268
7.14	Tangent Distance, Tangent Prop, and Manifold Tangent Classifier	270
8	Optimization for Training Deep Models	274
8.1	How Learning Differs from Pure Optimization	275
8.2	Challenges in Neural Network Optimization	282
8.3	Basic Algorithms	294
8.4	Parameter Initialization Strategies	301
8.5	Algorithms with Adaptive Learning Rates	306
8.6	Approximate Second-Order Methods	310
8.7	Optimization Strategies and Meta-Algorithms	317
9	Convolutional Networks	330
9.1	The Convolution Operation	331
9.2	Motivation	335
9.3	Pooling	339
9.4	Convolution and Pooling as an Infinitely Strong Prior	345
9.5	Variants of the Basic Convolution Function	347
9.6	Structured Outputs	358
9.7	Data Types	360
9.8	Efficient Convolution Algorithms	362
9.9	Random or Unsupervised Features	363