

3.2	Random Variables	56
3.3	Probability Distributions	56
3.4	Marginal Probability	58
3.5	Conditional Probability	59
3.6	The Chain Rule of Conditional Probabilities	59
3.7	Independence and Conditional Independence	60
3.8	Expectation, Variance and Covariance	60
3.9	Common Probability Distributions	62
3.10	Useful Properties of Common Functions	67
3.11	Bayes' Rule	70
3.12	Technical Details of Continuous Variables	71
3.13	Information Theory	73
3.14	Structured Probabilistic Models	75
4	Numerical Computation	80
4.1	Overflow and Underflow	80
4.2	Poor Conditioning	82
4.3	Gradient-Based Optimization	82
4.4	Constrained Optimization	93
4.5	Example: Linear Least Squares	96
5	Machine Learning Basics	98
5.1	Learning Algorithms	99
5.2	Capacity, Overfitting and Underfitting	110
5.3	Hyperparameters and Validation Sets	120
5.4	Estimators, Bias and Variance	122
5.5	Maximum Likelihood Estimation	131
5.6	Bayesian Statistics	135
5.7	Supervised Learning Algorithms	140
5.8	Unsupervised Learning Algorithms	146
5.9	Stochastic Gradient Descent	151
5.10	Building a Machine Learning Algorithm	153
5.11	Challenges Motivating Deep Learning	155
II	Deep Networks: Modern Practices	166
6	Deep Feedforward Networks	168
6.1	Example: Learning XOR	171
6.2	Gradient-Based Learning	177