

Contents

Preface	vii
1 Linear Algebra, Numerical Optimization, and Its Applications in Machine Learning	1
1.1 Scalars, Vectors, and Linear Functions	1
1.1.1 Scalars	1
1.1.2 Vectors	1
1.2 Linear Functions	4
1.3 Matrices	4
1.3.1 Transpose of a Matrix	4
1.3.2 Identity Matrix	4
1.3.3 Inverse of a Matrix	5
1.3.4 Representing Linear Equations in Matrix Form	5
1.4 Matrix Transformations	6
1.5 Norms	7
1.5.1 ℓ_2 Optimization	8
1.5.2 ℓ_1 Optimization	9
1.6 Rewriting the Regression Model in Matrix Notation	9
1.7 Cost of a n-Dimensional Function	10
1.8 Computing the Gradient of the Cost	11
1.8.1 Closed-Form Solution	11
1.8.2 Gradient Descent	12
1.9 An Example of Gradient Descent Optimization	13
1.10 Eigendecomposition	14
1.11 Singular Value Decomposition (SVD)	18
1.12 Principal Component Analysis (PCA)	21
1.12.1 PCA and SVD	22
1.13 Computational Errors	27
1.13.1 Rounding—Overflow and Underflow	28
1.13.2 Conditioning	28
1.14 Numerical Optimization	29

2	Probability and Distributions	31
2.1	Sources of Uncertainty	31
2.2	Random Experiment	32
2.3	Probability	32
2.3.1	Marginal Probability	33
2.3.2	Conditional Probability	34
2.3.3	The Chain Rule	34
2.4	Bayes' Rule	35
2.5	Probability Distribution	37
2.5.1	Discrete Probability Distribution	37
2.5.2	Continuous Probability Distribution	37
2.5.3	Cumulative Probability Distribution	37
2.5.4	Joint Probability Distribution	38
2.6	Measures of Central Tendency	38
2.7	Dispersion	39
2.8	Covariance and Correlation	39
2.9	Shape of a Distribution	41
2.10	Chebyshev's Inequality	41
2.11	Common Probability Distributions	42
2.11.1	Discrete Distributions	42
2.11.2	Continuous Distributions	43
2.11.3	Summary of Probability Distributions	45
2.12	Tests for Fit	46
2.12.1	Chi-Square Distribution	47
2.12.2	Chi-Square Test	48
2.13	Ratio Distributions	50
2.13.1	Student's t-Distribution	51
2.13.2	F-Distribution	54
3	Introduction to Machine Learning	57
3.1	Scientific Enquiry	58
3.1.1	Empirical Science	58
3.1.2	Theoretical Science	59
3.1.3	Computational Science	59
3.1.4	e-Science	59
3.2	Machine Learning	59
3.2.1	A Learning Task	60
3.2.2	The Performance Measure	60
3.2.3	The Experience	61
3.3	Train and Test Data	61
3.3.1	Training Error, Generalization (True) Error, and Test Error	61

3.4	Irreducible Error, Bias, and Variance	64
3.5	Bias–Variance Trade-off	66
3.6	Deriving the Expected Prediction Error	67
3.7	Underfitting and Overfitting	68
3.8	Regularization	69
3.9	Hyperparameters	71
3.10	Cross-Validation	72
3.11	Maximum Likelihood Estimation	72
3.12	Gradient Descent	75
3.13	Building a Machine Learning Algorithm	76
3.13.1	Challenges in Learning Algorithms	77
3.13.2	Curse of Dimensionality and Feature Engineering	77
3.14	Conclusion	78
4	Regression	79
4.1	Linear Regression	79
4.1.1	Hypothesis Function	79
4.1.2	Cost Function	80
4.2	Linear Regression as Ordinary Least Squares	81
4.3	Linear Regression as Maximum Likelihood	83
4.4	Gradient Descent	84
4.4.1	Gradient of RSS	84
4.4.2	Closed Form Solution	84
4.4.3	Step-by-Step Batch Gradient Descent	84
4.4.4	Writing the Batch Gradient Descent Application	85
4.4.5	Writing the Stochastic Gradient Descent Application	89
4.5	Linear Regression Assumptions	90
4.6	Summary of Regression Outputs	93
4.7	Ridge Regression	95
4.7.1	Computing the Gradient of Ridge Regression	97
4.7.2	Writing the Ridge Regression Gradient Descent Application	99
4.8	Assessing Performance	103
4.8.1	Sources of Error Revisited	104
4.8.2	Bias–Variance Trade-Off in Ridge Regression	106
4.9	Lasso Regression	107
4.9.1	Coordinate Descent for Least Squares Regression	108
4.9.2	Coordinate Descent for Lasso	109
4.9.3	Writing the Lasso Coordinate Descent Application	110
4.9.4	Implementing Coordinate Descent	112
4.9.5	Bias Variance Trade-Off in Lasso Regression	113

5	Classification	115
5.1	Linear Classifiers	115
5.1.1	Linear Classifier Model	116
5.1.2	Interpreting the Score	117
5.2	Logistic Regression	117
5.2.1	Likelihood Function	120
5.2.2	Model Selection with Log-Likelihood	120
5.2.3	Gradient Ascent to Find the Best Linear Classifier	121
5.2.4	Deriving the Log-Likelihood Function	122
5.2.5	Deriving the Gradient of Log-Likelihood	124
5.2.6	Gradient Ascent for Logistic Regression	125
5.2.7	Writing the Logistic Regression Application	125
5.2.8	A Comparison Using the BFGS Optimization Method	129
5.2.9	Regularization	131
5.2.10	ℓ_2 Regularized Logistic Regression	131
5.2.11	ℓ_2 Regularized Logistic Regression with Gradient Ascent	133
5.2.12	Writing the Ridge Logistic Regression with Gradient Ascent Application	133
5.2.13	Writing the Lasso Regularized Logistic Regression With Gradient Ascent Application	138
5.3	Decision Trees	143
5.3.1	Decision Tree Algorithm	145
5.3.2	Overfitting in Decision Trees	145
5.3.3	Control of Tree Parameters	146
5.3.4	Writing the Decision Tree Application	147
5.3.5	Unbalanced Data	152
5.4	Assessing Performance	153
5.4.1	Assessing Performance—Logistic Regression	155
5.5	Boosting	158
5.5.1	AdaBoost Learning Ensemble	160
5.5.2	AdaBoost: Learning from Weighted Data	160
5.5.3	AdaBoost: Updating the Weights	161
5.5.4	AdaBoost Algorithm	162
5.5.5	Writing the Weighted Decision Tree Algorithm	162
5.5.6	Writing the AdaBoost Application	168
5.5.7	Performance of our AdaBoost Algorithm	172
5.6	Other Variants	175
5.6.1	Bagging	175
5.6.2	Gradient Boosting	176
5.6.3	XGBoost	176

6 Clustering	179
6.1 The Clustering Algorithm	180
6.2 Clustering Algorithm as Coordinate Descent optimization	180
6.3 An Introduction to Text mining	181
6.3.1 Text Mining Application—Reading Multiple Text Files from Multiple Directories	181
6.3.2 Text Mining Application—Creating a Weighted tf-idf Document-Term Matrix	182
6.3.3 Text Mining Application—Exploratory Analysis	183
6.4 Writing the Clustering Application	183
6.4.1 Smart Initialization of k-means	193
6.4.2 Writing the <i>k</i> -means++ Application	193
6.4.3 Finding the Optimal Number of Centroids	199
6.5 Topic Modeling	201
6.5.1 Clustering and Topic Modeling	201
6.5.2 Latent Dirichlet Allocation for Topic Modeling	202
References and Further Reading	209



<http://www.springer.com/978-981-10-6807-2>

Machine Learning with R

Ghatak, A.

2017, XIX, 210 p. 56 illus., Hardcover

ISBN: 978-981-10-6807-2