

## MACHINE LEARNING

Marc Peter Deisenroth A. Aldo Faisal Cheng Soon Ong

## Contents

Forew	Foreword		
	Part I Mathematical Foundations	9	
1 1.1 1.2 1.3	<b>Introduction and Motivation</b> Finding Words for Intuitions Two Ways to Read This Book Exercises and Feedback	11 12 13 16	
2 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9	Linear Algebra Systems of Linear Equations Matrices Solving Systems of Linear Equations Vector Spaces Linear Independence Basis and Rank Linear Mappings Affine Spaces Further Reading Exercises	17 19 22 27 35 40 44 48 61 63 63	
<b>3</b> 3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 3.10	Analytic Geometry Norms Inner Products Lengths and Distances Angles and Orthogonality Orthonormal Basis Orthogonal Complement Inner Product of Functions Orthogonal Projections Rotations Further Reading Exercises	70 71 72 75 76 78 79 80 81 91 94 95	
<b>4</b> 4.1	Matrix Decompositions Determinant and Trace	98 99	

i

Draft (October 15, 2019) of "Mathematics for Machine Learning" ©2019 by M. P. Deisenroth, A. A. Faisal, and C. S. Ong. To be published by Cambridge University Press. https://mml-book.com.

ii		Contents
4.2	Eigenvalues and Eigenvectors	105
4.3	Cholesky Decomposition	114
4.4	Eigendecomposition and Diagonalization	115
4.5	Singular Value Decomposition	119
4.6	Matrix Approximation	129
4.7	Matrix Phylogeny	134
4.8	Further Reading	135
	Exercises	137
5	Vector Calculus	139
5.1	Differentiation of Univariate Functions	141
5.2	Partial Differentiation and Gradients	146
5.3	Gradients of Vector-Valued Functions	149
5.4	Gradients of Matrices	155
5.5	Useful Identities for Computing Gradients	158
5.6	Backpropagation and Automatic Differentiation	159
5.7	Higher-Order Derivatives	164
5.8	Linearization and Multivariate Taylor Series	165
5.9	Further Reading	170
	Exercises	170
6	Probability and Distributions	172
6.1	Construction of a Probability Space	172
6.2	Discrete and Continuous Probabilities	178
6.3	Sum Rule, Product Rule, and Bayes' Theorem	183
6.4	Summary Statistics and Independence	186
6.5	Gaussian Distribution	197
6.6	Conjugacy and the Exponential Family	205
6.7	Change of Variables/Inverse Transform	214
6.8	Further Reading	221
	Exercises	222
7	Continuous Optimization	225
7.1	Optimization Using Gradient Descent	227
7.2	Constrained Optimization and Lagrange Multipliers	233
7.3	Convex Optimization	236
7.4	Further Reading	246
	Exercises	247
	Part II Central Machine Learning Problems	249
8	When Models Meet Data	251
8.1	Data, Models, and Learning	251
8.2	Empirical Risk Minimization	258
8.3	Parameter Estimation	265
8.4	Probabilistic Modeling and Inference	272
8.5	Directed Graphical Models	278

Draft (2019-10-15) of "Mathematics for Machine Learning". Feedback: https://mml-book.com.

Contents		iii
8.6	Model Selection	283
9	Linear Regression	289
9.1	Problem Formulation	291
9.2	Parameter Estimation	292
9.3	Bayesian Linear Regression	303
9.4	Maximum Likelihood as Orthogonal Projection	313
9.5	Further Reading	315
10	Dimensionality Reduction with Principal Component Analysis	317
10.1	Problem Setting	318
10.2	Maximum Variance Perspective	320
10.3	Projection Perspective	325
10.4	Eigenvector Computation and Low-Rank Approximations	333
10.5	PCA in High Dimensions	335
10.6	Key Steps of PCA in Practice	336
10.7	Latent Variable Perspective	339
10.8	Further Reading	343
11	Density Estimation with Gaussian Mixture Models	348
11.1	Gaussian Mixture Model	349
11.2	Parameter Learning via Maximum Likelihood	350
11.3	EM Algorithm	360
11.4	Latent-Variable Perspective	363
11.5	Further Reading	368
12	Classification with Support Vector Machines	370
12.1	Separating Hyperplanes	372
12.2	Primal Support Vector Machine	374
12.3	Dual Support Vector Machine	383
12.4	Kernels	388
12.5	Numerical Solution	390
12.6	Further Reading	392
Refere	References	
Index	Index	