Machine Intelligence

Applied Mathematics and Machine Learning

Ruye Wang, Nick Richardson

Harvey Mudd College
Department of Engineering, Department of Mathematics

E-mail: rwang@hmc.edu, nrichardson@g.hmc.edu

Contents

I Fundamentals 2

1 Probabilities and Inference 2

2 Linear Algebra 2

3 Signal Processing 2

4 Optimization 2

II Representation 3

5 Smoothing, Compression, and Information Loss 3

6 Basic Basis Function Expansion 4

7 Adaptive Basis Function Methods 5

8 Latent Processes 5

9 Automatic Structure Discovery 5

10 Metric Learning 5

11 Representation examples 5

III Inference 6

12 Linear methods 6

13 Adaptive Basis Function Methods 6

IV Application 7

14 Step Zero 7

15 Exploratory Data Analysis 7

16 Model Selection 7

17 Hyperparameter Selection 7

18 Datasets and Distribution Shift 7

19 Application: Computer Vision 7

20 Application: Linguistics & Genomics 7

21 Application: Robotics & Control 7

V Frontiers 8

22 Model Composition 8

23 Returning to Control: Machine Learning in Society 8

24 Brain Machine Interfacing 8

25 Machine Intelligence: What will it take? 8

VI Appendices 8

Contents (Extendend)

I Fundamentals 2

1 Probabilities and Inference 2
1.1 Probability 2
1.2 Collections of Probabilities 2
1.2.1 Joint Probability 2
1.2.2 Marginal Probability 2
1.2.3 Conditional Probability 2
1.3 Independence 2
1.4 Product Rule & Sum Rule 2
1.5 Bayes' Theorem 2
1.6 Expectation 2
1.7 Variance & Covariance 2
1.8 Information & Entropy 2
1.9 Bounds on Signal Compression 2
1.10 Kullback-Leibler Divergence 2
1.11 Inductive Bias 2
1.12 The Language of Inference 2
1.12.1 Prior 2
1.12.2 Likelihood 2
1.12.3 Evidence/Marginal Likelihood 2
1.13 Point Estimation 2
1.13.1 Maximum Likelihood 2
1.13.2 Maximum a Posteriori 2
1.14 Laplace's Method 2
1.15 Variational Inference 2
1.16 Markov Chain Monte Carlo 2
1.16.1 Importance Sampling 2
1.16.2 Rejection Sampling 2
1.16.3 Hamiltonian MCMC 2

2 Linear Algebra 2
2.1 Vectors 2
2.1.1 Vector addition and scaling 2
2.1.2 Inner Products 2
2.1.3 Vector Norms 2
2.1.4 Angle 2
2.1.5 Vector Mean and Variance 2
2.1.6 Linearity and Approximations from Calculus 2

2.1.7 Linear Independence and Basis 2
2.1.8 Orthogonality 2
2.1.9 Vector Mean and Variance 2
2.1.10 Vectorized data 2
2.1.11 k-means clustering: A Taste of Automatic Structure Discovery 2
2.2 Matrices 2
2.2.1 Matrix addition and transpose 2
2.2.2 Matrix-vector and matrix-matrix multiplication 2
2.2.3 Geometric matrices 2
2.2.4 Convolution matrices 2
2.2.5 Finite Differences 2
2.2.6 Matrix representations of graphs 2
2.2.7 Affine functions 2
2.2.8 Systems of Linear Equations & Compositions of Linear Functions 2
2.2.9 Matrix Inverses & Psuedo-inverses 2
2.2.10 Matrix Powers & Eigensystems 2
2.2.11 Common matrix factorizations 2
2.2.12 Singular Value Decomposition 2
2.2.13 Least Squares 2

3 Signal Processing **2**
3.1 Basic Processing 2
3.1.1 Rescaling and centering 2
3.1.2 Histograms 2
3.1.3 Spectral Decay and effective dimensionality 2
3.1.4 Moving Functionals 2
3.1.5 Interpolation and Resampling 2
3.1.6 Linear Filters 2
3.1.7 Nonlinear Filters 2
3.1.8 Fourier Transform 2
3.1.9 Wavelet Transforms 2
3.1.10 Shapelets 2
3.2 Signal Inference 2
3.2.1 Kalman Filter 2
3.2.2 Hidden Markov model 2

4 Optimization **2**

II Representation **3**

5 Smoothing, Compression, and Information Loss 3
5.0.1 Basis sparsity and the Fourier Transform 3
5.0.2 Geometric Compression 3
5.0.3 Introduction to Dimensionality Reduction 3

6 Basic Basis Function Expansion 4
6.1 Polynomial Basis 4
6.2 Periodic Basis 4
6.3 Piecewise Linear 4
6.4 Radial Basis Functions 4
6.5 Sigmoid Functions 4

7 Adaptive Basis Function Methods 5
7.1 Kernels 5
7.1.1 The Gram matrix 5
7.1.2 The Kernel Trick 5
7.1.3 Learning the Kernel 5
7.2 Gaussian Processes 5
7.3 Neural Networks 5
7.4 Kernel Machines 5

8 Latent Processes 5
8.1 Discrete Latent Processes 5
8.2 Continuous Latent Processes 5
8.3 Latent Processes as Probabilistic Graphical Models 5

9 Automatic Structure Discovery 5
9.1 Clustering and Graphical Community Identification 5
9.2 Hierarchical clustering 5
9.3 Mean and mediod based clustering 5
9.4 Spectral clustering 5
9.5 Principal components analysis 5
9.6 Linear discriminant analysis 5
9.7 T-distributed stochastic neighbor embedding 5
9.8 Laplacian Eigenmaps and UMAP 5

10 Metric Learning 5

11 Representation examples 5
11.1 Image and video 5
11.2 Speech and sound 5
11.3 Text and genetic sequence 5

III Inference 6

12 Linear methods 6
12.1 Linear Regression 6
12.1.1 Least Squares and Linear Regression 6
12.1.2 Basis Function Expansion 6
12.1.3 Solution Penalties: A Return to Vector Norms and Geometry 6
12.1.4 Bayesian Linear Regression 6
12.2 Logistic Regression 6
12.3 Generalized Linear Models 6
12.3.1 The Exponential Family 6
12.3.2 Conjugacy and Priors 6

13 Adaptive Basis Function Methods 6
13.1 Inference with Neural Networks 6
13.1.1 Single Neurons 6
13.1.2 Feedforward Neural Networks 6
13.1.3 Structured Neural Networks 6
13.1.4 Weight Intialization and Random Matrices 6
13.1.5 Optimization in Neural Networks 6
13.1.6 Probabilistic Neural Networks 6
13.2 Inference with Kernel Machines 6
13.3 Inference with Gaussian Processes 6

IV Application 7

14 Step Zero 7
14.1 Understanding the Problem You Care About 7
14.2 Specifying the Inductive Bias 7

15 Exploratory Data Analysis 7
15.1 Dataset Versioning 7
15.2 Visualization 7
15.3 Guaging Problem Difficulty 7
15.4 Starting Simply 7

16 Model Selection 7
16.1 Standard Model Selection 7
16.2 Bayesian Model Selection 7

17 Hyperparameter Selection 7
17.1 Heuristics 7
17.2 Bayesian Optimization 7

18 Datasets and Distribution Shift 7
18.1 Training on the Test Set 7
18.2 Internal Covariate Shift 7
18.3 Distribution Shift 7

19 Application: Computer Vision 7

20 Application: Linguistics & Genomics 7

21 Application: Robotics & Control 7

V Frontiers 8

22 Model Composition 8
22.1 Probabilistic Graphics Models and Adaptive Basis Methods 8

23 Returning to Control: Machine Learning in Society 8
23.1 Autonomous Vehicles 8
23.2 Autonomy in Weapon Systems 8
23.3 Robustness in Machine Learning 8

24 Brain Machine Interfacing 8
24.1 Studying the Brain 8
24.2 Imitating the Brain 8
24.3 Machine intelligence as analogies for neurocomputation 8

25 Machine Intelligence: What will it take? 8
25.1 The AI effect 8
25.2 Hype and Reality 8
25.3 Constructing Cockroaches 8

VI Appendices 8
25.4 Notation 8
25.5 Calculus 8
25.5.1 Derivatives 8
25.5.2 Chain Rule & Product Rule 8
25.5.3 Integrals 8
25.6 Complex numbers as dynamic objects 8

PART

I

Fundamentals

1 Probabilities and Inference
1.1 Probability
1.2 Collections of Probabilities
1.2.1 Joint Probability
1.2.2 Marginal Probability
1.2.3 Conditional Probability
1.3 Independence
1.4 Product Rule & Sum Rule
1.5 Bayes' Theorem
1.6 Expectation
1.7 Variance & Covariance
1.8 Information & Entropy
1.9 Bounds on Signal Compression
1.10 Kullback-Leibler Divergence
1.11 Inductive Bias
1.12 The Language of Inference
1.12.1 Prior
1.12.2 Likelihood
1.12.3 Evidence/Marginal Likelihood
1.13 Point Estimation
1.13.1 Maximum Likelihood
1.13.2 Maximum a Posteriori
1.14 Laplace's Method
1.15 Variational Inference
1.16 Markov Chain Monte Carlo
1.16.1 Importance Sampling
1.16.2 Rejection Sampling
1.16.3 Hamiltonian MCMC
2 Linear Algebra
2.1 Vectors
2.1.1 Vector addition and scaling
2.1.2 Inner Products
2.1.3 Vector Norms
2.1.4 Angle
2.1.5 Vector Mean and Variance
2.1.6 Linearity and Approximations from Calculus

PART

II

Representation

5 Smoothing, Compression, and Information Loss

5.0.1 Basis sparsity and the Fourier Transform

5.0.2 Geometric Compression

5.0.3 Introduction to Dimensionality Reduction

6 Basic Basis Function Expansion

6.1 Polynomial Basis

6.2 Periodic Basis

6.3 Piecewise Linear

6.4 Radial Basis Functions

6.5 Sigmoid Functions

7 Adaptive Basis Function Methods

7.1 Kernels

7.1.1 The Gram matrix

7.1.2 The Kernel Trick

7.1.3 Learning the Kernel

7.2 Gaussian Processes

7.3 Neural Networks

7.4 Kernel Machines

8 Latent Processes

8.1 Discrete Latent Processes

8.2 Continuous Latent Processes

8.3 Latent Processes as Probabilistic Graphical Models

9 Automatic Structure Discovery

9.1 Clustering and Graphical Community Identification

9.2 Hierarchical clustering

9.3 Mean and mediod based clustering

9.4 Spectral clustering

9.5 Principal components analysis

9.6 Linear discriminant analysis

9.7 T-distributed stochastic neighbor embedding

9.8 Laplacian Eigenmaps and UMAP

10 Metric Learning

11 Representation examples

11.1 Image and video

11.2 Speech and sound

11.3 Text and genetic sequence

PART

III

Inference

12 Linear methods

12.1 Linear Regression

12.1.1 Least Squares and Linear Regression

12.1.2 Basis Function Expansion

12.1.3 Solution Penalties: A Return to Vector Norms and Geometry

12.1.4 Bayesian Linear Regression

12.2 Logistic Regression

12.3 Generalized Linear Models

12.3.1 The Exponential Family

12.3.2 Conjugacy and Priors

13 Adaptive Basis Function Methods

13.1 Inference with Neural Networks

13.1.1 Single Neurons

13.1.2 Feedforward Neural Networks

13.1.3 Structured Neural Networks

13.1.4 Weight Intialization and Random Matrices

13.1.5 Optimization in Neural Networks

13.1.6 Probabilistic Neural Networks

13.2 Inference with Kernel Machines

13.3 Inference with Gaussian Processes

PART

IV

Application

14 Step Zero

14.1 Understanding the Problem You Care About

14.2 Specifying the Inductive Bias

15 Exploratory Data Analysis

15.1 Dataset Versioning

15.2 Visualization

15.3 Guaging Problem Difficulty

15.4 Starting Simply

16 Model Selection

16.1 Standard Model Selection

16.2 Bayesian Model Selection

17 Hyperparameter Selection

17.1 Heuristics

17.2 Bayesian Optimization

18 Datasets and Distribution Shift

18.1 Training on the Test Set

18.2 Internal Covariate Shift

18.3 Distribution Shift

19 Application: Computer Vision

20 Application: Linguistics & Genomics

21 Application: Robotics & Control

PART
V

Frontiers

22 Model Composition

22.1 Probabilistic Graphics Models and Adaptive Basis Methods

23 Returning to Control: Machine Learning in Society

23.1 Autonomous Vehicles

23.2 Autonomy in Weapon Systems

23.3 Robustness in Machine Learning

24 Brain Machine Interfacing

24.1 Studying the Brain

24.2 Imitating the Brain

24.3 Machine intelligence as analogies for neurocomputation

25 Machine Intelligence: What will it take?

25.1 The AI effect

25.2 Hype and Reality

25.3 Constructing Cockroaches

PART
VI

Appendices

25.4 Notation

25.5 Calculus

25.5.1 Derivatives

25.5.2 Chain Rule & Product Rule

25.5.3 Integrals

25.6 Complex numbers as dynamic objects