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

II Representation 3

4 Smoothing, Compression, and Information Loss 3

5 Basic Basis Function Expansion 4

6 Adaptive Basis Function Methods 5

7 Latent Processes 5

8 Automatic Structure Discovery 5

9 Metric Learning 5

10 Representation examples 5

III Inference 6

11 Linear methods 6

12 Adaptive Basis Function Methods 6

IV Application 7

13 Step Zero 7

14 Exploratory Data Analysis 7

15 Model Selection 7

16 Hyperparameter Selection 7

17 Datasets and Distribution Shift 7

18 Application: Computer Vision 7

19 Application: Linguistics & Genomics 7

20 Application: Robotics & Control 7

V Frontiers 8

21 Model Composition 8

22 Returning to Control: Machine Learning in Society 8

23 Brain Machine Interfacing 8

24 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

II Representation 3

4 Smoothing, Compression, and Information Loss 3
4.0.1 Basis sparsity and the Fourier Transform 3
4.0.2 Geometric Compression 3
4.0.3 Introduction to Dimensionality Reduction 3

5 Basic Basis Function Expansion **4**
5.1 Polynomial Basis 4
5.2 Periodic Basis 4
5.3 Piecewise Linear 4
5.4 Radial Basis Functions 4
5.5 Sigmoid Functions 4

6 Adaptive Basis Function Methods **5**
6.1 Kernels 5
6.1.1 The Gram matrix 5
6.1.2 The Kernel Trick 5
6.1.3 Learning the Kernel 5
6.2 Gaussian Processes 5
6.3 Neural Networks 5
6.4 Kernel Machines 5

7 Latent Processes **5**
7.1 Discrete Latent Processes 5
7.2 Continuous Latent Processes 5
7.3 Latent Processes as Probabilistic Graphical Models 5

8 Automatic Structure Discovery **5**
8.1 Clustering and Graphical Community Identification 5
8.2 Hierarchical clustering 5
8.3 Mean and mediod based clustering 5
8.4 Spectral clustering 5
8.5 Principal components analysis 5
8.6 Linear discriminant analysis 5
8.7 T-distributed stochastic neighbor embedding 5
8.8 Laplacian Eigenmaps and UMAP 5

9 Metric Learning **5**

10 Representation examples **5**
10.1 Image and video 5
10.2 Speech and sound 5
10.3 Text and genetic sequence 5

III Inference **6**

11 Linear methods **6**
11.1 Linear Regression 6
11.1.1 Least Squares and Linear Regression 6
11.1.2 Basis Function Expansion 6

11.1.3 Solution Penalties: A Return to Vector Norms and Geometry 6
11.1.4 Bayesian Linear Regression 6
11.2 Logistic Regression 6
11.3 Generalized Linear Models 6
11.3.1 The Exponential Family 6
11.3.2 Conjugacy and Priors 6

12 Adaptive Basis Function Methods 6
12.1 Inference with Neural Networks 6
12.1.1 Single Neurons 6
12.1.2 Feedforward Neural Networks 6
12.1.3 Structured Neural Networks 6
12.1.4 Weight Intialization and Random Matrices 6
12.1.5 Optimization in Neural Networks 6
12.1.6 Probabilistic Neural Networks 6
12.2 Inference with Kernel Machines 6
12.3 Inference with Gaussian Processes 6

IV Application 7

13 Step Zero 7
13.1 Understanding the Problem You Care About 7
13.2 Specifying the Inductive Bias 7

14 Exploratory Data Analysis 7
14.1 Dataset Versioning 7
14.2 Visualization 7
14.3 Guaging Problem Difficulty 7
14.4 Starting Simply 7

15 Model Selection 7
15.1 Standard Model Selection 7
15.2 Bayesian Model Selection 7

16 Hyperparameter Selection 7
16.1 Heuristics 7
16.2 Bayesian Optimization 7

17 Datasets and Distribution Shift 7
17.1 Training on the Test Set 7
17.2 Internal Covariate Shift 7
17.3 Distribution Shift 7

18 Application: Computer Vision 7

19 Application: Linguistics & Genomics 7

20 Application: Robotics & Control 7

V Frontiers 8

21 Model Composition 8

21.1 Probabilistic Graphics Models and Adaptive Basis Methods 8

22 Returning to Control: Machine Learning in Society 8

22.1 Autonomous Vehicles 8

22.2 Autonomy in Weapon Systems 8

22.3 Robustness in Machine Learning 8

23 Brain Machine Interfacing 8

23.1 Studying the Brain 8

23.2 Imitating the Brain 8

23.3 Machine intelligence as analogies for neurocomputation 8

24 Machine Intelligence: What will it take? 8

24.1 The AI effect 8

24.2 Hype and Reality 8

24.3 Constructing Cockroaches 8

VI Appendices 8

24.4 Notation 8

24.5 Calculus 8

24.5.1 Derivatives 8

24.5.2 Chain Rule & Product Rule 8

24.5.3 Integrals 8

24.6 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

4 Smoothing, Compression, and Information Loss

4.0.1 Basis sparsity and the Fourier Transform

4.0.2 Geometric Compression

4.0.3 Introduction to Dimensionality Reduction

5 Basic Basis Function Expansion

5.1 Polynomial Basis

5.2 Periodic Basis

5.3 Piecewise Linear

5.4 Radial Basis Functions

5.5 Sigmoid Functions

6 Adaptive Basis Function Methods

6.1 Kernels

6.1.1 The Gram matrix

6.1.2 The Kernel Trick

6.1.3 Learning the Kernel

6.2 Gaussian Processes

6.3 Neural Networks

6.4 Kernel Machines

7 Latent Processes

7.1 Discrete Latent Processes

7.2 Continuous Latent Processes

7.3 Latent Processes as Probabilistic Graphical Models

8 Automatic Structure Discovery

8.1 Clustering and Graphical Community Identification

8.2 Hierarchical clustering

8.3 Mean and mediod based clustering

8.4 Spectral clustering

8.5 Principal components analysis

8.6 Linear discriminant analysis

8.7 T-distributed stochastic neighbor embedding

8.8 Laplacian Eigenmaps and UMAP

9 Metric Learning

10 Representation examples

10.1 Image and video

10.2 Speech and sound

10.3 Text and genetic sequence

PART

III

Inference

11 Linear methods

11.1 Linear Regression

11.1.1 Least Squares and Linear Regression

11.1.2 Basis Function Expansion

11.1.3 Solution Penalties: A Return to Vector Norms and Geometry

11.1.4 Bayesian Linear Regression

11.2 Logistic Regression

11.3 Generalized Linear Models

11.3.1 The Exponential Family

11.3.2 Conjugacy and Priors

12 Adaptive Basis Function Methods

12.1 Inference with Neural Networks

12.1.1 Single Neurons

12.1.2 Feedforward Neural Networks

12.1.3 Structured Neural Networks

12.1.4 Weight Intialization and Random Matrices

12.1.5 Optimization in Neural Networks

12.1.6 Probabilistic Neural Networks

12.2 Inference with Kernel Machines

12.3 Inference with Gaussian Processes

PART

IV

Application

13 Step Zero

13.1 Understanding the Problem You Care About

13.2 Specifying the Inductive Bias

14 Exploratory Data Analysis

14.1 Dataset Versioning

14.2 Visualization

14.3 Guaging Problem Difficulty

14.4 Starting Simply

15 Model Selection

15.1 Standard Model Selection

15.2 Bayesian Model Selection

16 Hyperparameter Selection

16.1 Heuristics

16.2 Bayesian Optimization

17 Datasets and Distribution Shift

17.1 Training on the Test Set

17.2 Internal Covariate Shift

17.3 Distribution Shift

18 Application: Computer Vision

19 Application: Linguistics & Genomics

20 Application: Robotics & Control

PART V

Frontiers

21 Model Composition

21.1 Probabilistic Graphics Models and Adaptive Basis Methods

22 Returning to Control: Machine Learning in Society

22.1 Autonomous Vehicles

22.2 Autonomy in Weapon Systems

22.3 Robustness in Machine Learning

23 Brain Machine Interfacing

23.1 Studying the Brain

23.2 Imitating the Brain

23.3 Machine intelligence as analogies for neurocomputation

24 Machine Intelligence: What will it take?

24.1 The AI effect

24.2 Hype and Reality

24.3 Constructing Cockroaches

PART VI

Appendices

24.4 Notation

24.5 Calculus

24.5.1 Derivatives

24.5.2 Chain Rule & Product Rule

24.5.3 Integrals

24.6 Complex numbers as dynamic objects