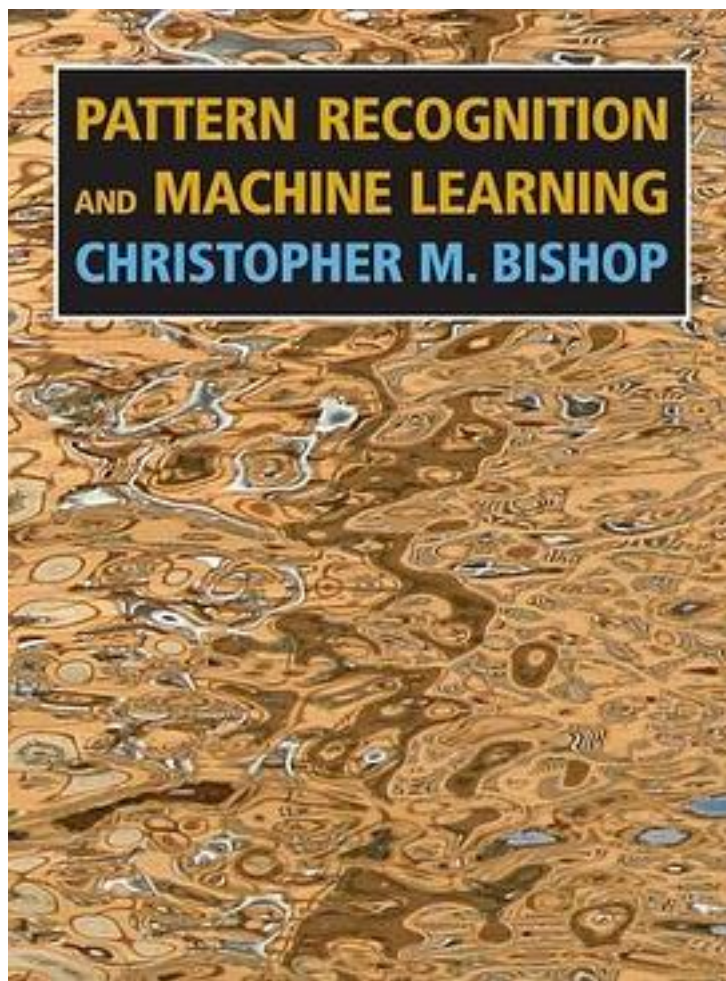


Pattern Recognition and Machine Learning



[Pattern Recognition and Machine Learning 下载链接1](#)

著者:Christopher M. Bishop

出版者:Springer

出版时间:2016-8-23

装帧:Paperback

isbn:9781493938438

The book is suitable for courses on machine learning, statistics, computer science, signal processing, computer vision, data mining, and bioinformatics. Extensive support

is provided for course instructors, including more than 400 exercises, graded according to difficulty. Example solutions for a subset of the exercises are available from the book web site, while solutions for the remainder can be obtained by instructors from the publisher. The book is supported by a great deal of additional material, and the reader is encouraged to visit the book web site for the latest information.

作者介绍:

Christopher M. Bishop is Deputy Director of Microsoft Research Cambridge, and holds a Chair in Computer Science at the University of Edinburgh. He is a Fellow of Darwin College Cambridge, a Fellow of the Royal Academy of Engineering, and a Fellow of the Royal Society of Edinburgh. His previous textbook "Neural Networks for Pattern Recognition" has been widely adopted.

目录: 1 Introduction 1

1.1 Example: Polynomial Curve Fitting	4
1.2 Probability Theory	12
1.2.1 Probability densities	17
1.2.2 Expectations and covariances	19
1.2.3 Bayesian probabilities	21
1.2.4 The Gaussian distribution	24
1.2.5 Curve fitting re-visited	28
1.2.6 Bayesian curve fitting	30
1.3 Model Selection	32
1.4 The Curse of Dimensionality	33
1.5 Decision Theory	38
1.5.1 Minimizing the misclassification rate	39
1.5.2 Minimizing the expected loss	41
1.5.3 The reject option	42
1.5.4 Inference and decision	42
1.5.5 Loss functions for regression	46
1.6 Information Theory	48
1.6.1 Relative entropy and mutual information	55
Exercises	58
2 Probability Distributions 67	
2.1 Binary Variables	68
2.1.1 The beta distribution	71
2.2 Multinomial Variables	74
2.2.1 The Dirichlet distribution	76
2.3 The Gaussian Distribution	78
2.3.1 Conditional Gaussian distributions	85
2.3.2 Marginal Gaussian distributions	88
2.3.3 Bayes' theorem for Gaussian variables	90
2.3.4 Maximum likelihood for the Gaussian	93
2.3.5 Sequential estimation	94
2.3.6 Bayesian inference for the Gaussian	97
2.3.7 Student's t-distribution	102
2.3.8 Periodic variables	105
2.3.9 Mixtures of Gaussians	110
2.4 The Exponential Family	113
2.4.1 Maximum likelihood and sufficient statistics	116
2.4.2 Conjugate priors	117

2.4.3 Noninformative priors	117
2.5 Nonparametric Methods	120
2.5.1 Kernel density estimators	122
2.5.2 Nearest-neighbour methods	124
Exercises	127
3 Linear Models for Regression	137
3.1 Linear Basis Function Models	138
3.1.1 Maximum likelihood and least squares	140
3.1.2 Geometry of least squares	143
3.1.3 Sequential learning	143
3.1.4 Regularized least squares	144
3.1.5 Multiple outputs	146
3.2 The Bias-Variance Decomposition	147
3.3 Bayesian Linear Regression	152
3.3.1 Parameter distribution	153
3.3.2 Predictive distribution	156
3.3.3 Equivalent kernel	157
3.4 Bayesian Model Comparison	161
3.5 The Evidence Approximation	165
3.5.1 Evaluation of the evidence function	166
3.5.2 Maximizing the evidence function	168
3.5.3 Effective number of parameters	170
3.6 Limitations of Fixed Basis Functions	172
Exercises	173
4 Linear Models for Classification	179
4.1 Discriminant Functions	181
4.1.1 Two classes	181
4.1.2 Multiple classes	182
4.1.3 Least squares for classification	184
4.1.4 Fisher's linear discriminant	186
4.1.5 Relation to least squares	189
4.1.6 Fisher's discriminant for multiple classes	191
4.1.7 The perceptron algorithm	192
4.2 Probabilistic Generative Models	196
4.2.1 Continuous inputs	198
4.2.2 Maximum likelihood solution	200
4.2.3 Discrete features	202
4.2.4 Exponential family	202
4.3 Probabilistic Discriminative Models	203
4.3.1 Fixed basis functions	204
4.3.2 Logistic regression	205
4.3.3 Iterative reweighted least squares	207
4.3.4 Multiclass logistic regression	209
4.3.5 Probit regression	210
4.3.6 Canonical link functions	212
4.4 The Laplace Approximation	213
4.4.1 Model comparison and BIC	216
4.5 Bayesian Logistic Regression	217
4.5.1 Laplace approximation	217
4.5.2 Predictive distribution	218
Exercises	220
5 Neural Networks	225
5.1 Feed-forward Network Functions	227
5.1.1 Weight-space symmetries	231

5.2 Network Training	232
5.2.1 Parameter optimization	236
5.2.2 Local quadratic approximation	237
5.2.3 Use of gradient information	239
5.2.4 Gradient descent optimization	240
5.3 Error Backpropagation	241
5.3.1 Evaluation of error-function derivatives	242
5.3.2 A simple example	245
5.3.3 Efficiency of backpropagation	246
5.3.4 The Jacobian matrix	247
5.4 The Hessian Matrix	249
5.4.1 Diagonal approximation	250
5.4.2 Outer product approximation	251
5.4.3 Inverse Hessian	252
5.4.4 Finite differences	252
5.4.5 Exact evaluation of the Hessian	253
5.4.6 Fast multiplication by the Hessian	254
5.5 Regularization in Neural Networks	256
5.5.1 Consistent Gaussian priors	257
5.5.2 Early stopping	259
5.5.3 Invariances	261
5.5.4 Tangent propagation	263
5.5.5 Training with transformed data	265
5.5.6 Convolutional networks	267
5.5.7 Soft weight sharing	269
5.6 Mixture Density Networks	272
5.7 Bayesian Neural Networks	277
5.7.1 Posterior parameter distribution	278
5.7.2 Hyperparameter optimization	280
5.7.3 Bayesian neural networks for classification	281
Exercises	284
6 Kernel Methods	291
6.1 Dual Representations	293
6.2 Constructing Kernels	294
6.3 Radial Basis Function Networks	299
6.3.1 Nadaraya-Watson model	301
6.4 Gaussian Processes	303
6.4.1 Linear regression revisited	304
6.4.2 Gaussian processes for regression	306
6.4.3 Learning the hyperparameters	311
6.4.4 Automatic relevance determination	312
6.4.5 Gaussian processes for classification	313
6.4.6 Laplace approximation	315
6.4.7 Connection to neural networks	319
Exercises	320
7 Sparse Kernel Machines	325
7.1 Maximum Margin Classifiers	326
7.1.1 Overlapping class distributions	331
7.1.2 Relation to logistic regression	336
7.1.3 Multiclass SVMs	338
7.1.4 SVMs for regression	339
7.1.5 Computational learning theory	344
7.2 Relevance Vector Machines	345
7.2.1 RVM for regression	345

7.2.2 Analysis of sparsity	349
7.2.3 RVM for classification	353
Exercises	357
8 Graphical Models	359
8.1 Bayesian Networks	360
8.1.1 Example: Polynomial regression	362
8.1.2 Generative models	365
8.1.3 Discrete variables	366
8.1.4 Linear-Gaussian models	370
8.2 Conditional Independence	372
8.2.1 Three example graphs	373
8.2.2 D-separation	378
8.3 Markov Random Fields	383
8.3.1 Conditional independence properties	383
8.3.2 Factorization properties	384
8.3.3 Illustration: Image de-noising	387
8.3.4 Relation to directed graphs	390
8.4 Inference in Graphical Models	393
8.4.1 Inference on a chain	394
8.4.2 Trees	398
8.4.3 Factor graphs	399
8.4.4 The sum-product algorithm	402
8.4.5 The max-sum algorithm	411
8.4.6 Exact inference in general graphs	416
8.4.7 Loopy belief propagation	417
8.4.8 Learning the graph structure	418
Exercises	418
9 Mixture Models and EM	423
9.1 K-means Clustering	424
9.1.1 Image segmentation and compression	428
9.2 Mixtures of Gaussians	430
9.2.1 Maximum likelihood	432
9.2.2 EM for Gaussian mixtures	435
9.3 An Alternative View of EM	439
9.3.1 Gaussian mixtures revisited	441
9.3.2 Relation to K-means	443
9.3.3 Mixtures of Bernoulli distributions	444
9.3.4 EM for Bayesian linear regression	448
9.4 The EM Algorithm in General	450
Exercises	455
10 Approximate Inference	461
10.1 Variational Inference	462
10.1.1 Factorized distributions	464
10.1.2 Properties of factorized approximations	466
10.1.3 Example: The univariate Gaussian	470
10.1.4 Model comparison	473
10.2 Illustration: Variational Mixture of Gaussians	474
10.2.1 Variational distribution	475
10.2.2 Variational lower bound	481
10.2.3 Predictive density	482
10.2.4 Determining the number of components	483
10.2.5 Induced factorizations	485
10.3 Variational Linear Regression	486
10.3.1 Variational distribution	486

10.3.2 Predictive distribution	488
10.3.3 Lower bound	489
10.4 Exponential Family Distributions	490
10.4.1 Variational message passing	491
10.5 Local Variational Methods	493
10.6 Variational Logistic Regression	498
10.6.1 Variational posterior distribution	498
10.6.2 Optimizing the variational parameters	500
10.6.3 Inference of hyperparameters	502
10.7 Expectation Propagation	505
10.7.1 Example: The clutter problem	511
10.7.2 Expectation propagation on graphs	513
Exercises	517
11 Sampling Methods	523
11.1 Basic Sampling Algorithms	526
11.1.1 Standard distributions	526
11.1.2 Rejection sampling	528
11.1.3 Adaptive rejection sampling	530
11.1.4 Importance sampling	532
11.1.5 Sampling-importance-resampling	534
11.1.6 Sampling and the EM algorithm	536
11.2 Markov Chain Monte Carlo	537
11.2.1 Markov chains	539
11.2.2 The Metropolis-Hastings algorithm	541
11.3 Gibbs Sampling	542
11.4 Slice Sampling	546
11.5 The Hybrid Monte Carlo Algorithm	548
11.5.1 Dynamical systems	548
11.5.2 Hybrid Monte Carlo	552
11.6 Estimating the Partition Function	554
Exercises	556
12 Continuous Latent Variables	559
12.1 Principal Component Analysis	561
12.1.1 Maximum variance formulation	561
12.1.2 Minimum-error formulation	563
12.1.3 Applications of PCA	565
12.1.4 PCA for high-dimensional data	569
12.2 Probabilistic PCA	570
12.2.1 Maximum likelihood PCA	574
12.2.2 EM algorithm for PCA	577
12.2.3 Bayesian PCA	580
12.2.4 Factor analysis	583
12.3 Kernel PCA	586
12.4 Nonlinear Latent Variable Models	591
12.4.1 Independent component analysis	591
12.4.2 Autoassociative neural networks	592
12.4.3 Modelling nonlinear manifolds	595
Exercises	599
13 Sequential Data	605
13.1 Markov Models	607
13.2 Hidden Markov Models	610
13.2.1 Maximum likelihood for the HMM	615
13.2.2 The forward-backward algorithm	618
13.2.3 The sum-product algorithm for the HMM	625

13.2.4 Scaling factors	627
13.2.5 The Viterbi algorithm	629
13.2.6 Extensions of the hidden Markov model	631
13.3 Linear Dynamical Systems	635
13.3.1 Inference in LDS	638
13.3.2 Learning in LDS	642
13.3.3 Extensions of LDS	644
13.3.4 Particle filters	645
Exercises	646
14 Combining Models	653
14.1 Bayesian Model Averaging	654
14.2 Committees	655
14.3 Boosting	657
14.3.1 Minimizing exponential error	659
14.3.2 Error functions for boosting	661
14.4 Tree-based Models	663
14.5 Conditional Mixture Models	666
14.5.1 Mixtures of linear regression models	667
14.5.2 Mixtures of logistic models	670
14.5.3 Mixtures of experts	672
Exercises	674
Appendix A Data Sets	677
Appendix B Probability Distributions	685
Appendix C Properties of Matrices	695
Appendix D Calculus of Variations	703
Appendix E Lagrange Multipliers	707
References	711
• • • • •	(收起)

[Pattern Recognition and Machine Learning_ 下载链接1](#)

标签

机器学习

人工智能

计算机

管理

求购有二手的pattern

有电子版

技术

成长

评论

这本书06年出了一版 到了17年又出一版 对比时间跨度长达十年的两个版本，感觉基于统计的机器学习进展不是那么疯狂 == 不像RDL 基本就不能看书了 看完wiki补一下基础 就要直接看论文了

[Pattern Recognition and Machine Learning_下载链接1](#)

书评

两年多以前有个Machine Learning课以PRML为参考书，当时就觉得这书相当的好。可惜一直以来没认真读完。最近稍闲终于重新读了一遍，比较有收获。这书给人的最大的印象可能是everything has a Bayesian version或者说everything can be Bayesianized，比如PRML至少给出了以下Bay...

这几天没事把尾巴扫了。
如果想做ML无论是theory（tcusers请先别吐槽好吧，以后会有槽吐你们的）、algorithm还是application此书都是必读，而且书只读这一本足够了。ML吹破天还是那点内容，想学“fashion”的concept有那么多paper、review，看书是自取其辱。有人说此书遗憾没有...

我是学工程的，读过很多统计，模式识别，数据挖掘的书。比如Andrew Gelman 的 Bayesian data Analysis； Trevor Hastie 的 The Elements of Statistical Learning等等。。。我发现一个问题，但凡是统计系人出的书，我读起来都特别困难，比如以上提到的两本，基本读到第四第...

这两天因为读文章的需要，重新翻了翻这本书。觉得@raullew在<http://book.douban.com/review/4474434/>中提到的问题的确是这本书的一个缺陷。是否真正了解一个东西，不仅取决于你是否了解这个东西的特性，还取决于你能不能把它和相似的东西区分开。比如说，你要学习什么是猫，...

在Bishop的这本PRML之前，学习machine learning的标准教材一般是Tom Mitchell的machine learning以及Duda&Hart的Pattern Classification (那个年代ML与PR非常大的重合之处)。不可否认，这两本书都是ML领域的经典教材，但是由于成书时间太早，基本上都属于上古读物，已经不大适...

这本书的独到之处就是Bishop能够将看似毫无联系的方法统一在一个完整的框架下。虽然@raullew在<http://book.douban.com/review/4474434/>吐槽，但这正是Bishop想要传递的这本书的精髓所在。如果仅仅是各个算法的单独罗列，那我觉得去看wikipedia好了，还是免费的。相比之下，Du...

断断续续看到现在大概完成了前11章，其间收集了一些资料，书评等完整看过之后再补上。PRML的数学不是很大问题，因为很多用到的技巧都给出了(大量出现在第2章，少量出现在第8章)，或者是以附注的形式添加到了习题中，而习题是有答案的。主要障碍是书中的错误很多，有英文版错...

个人认为这是机器学习领域必读的一本书，甚至是目前最好的书。但这本书太过于Bayesian, 作者对任何算法都试图从概率和Bayesian的角度来进行解释。这本书不适合作为第一本教材，因为其为了将书中内容串联起来，忽视了这些内容的本来面貌，我印象比较深刻的地方有：第1.2.5节...

我是一名研一的学生，方向不是机器学习方向，但是对这方面很感兴趣。看过一篇blog说，当下所说的机器学习其实分两种，一种如本书，可称为统计机器学习，另外一种是人工智能领域，这两种有交叉，但是研究内容有很大不同。初读这书，刚觉得很罗嗦，加上是英语，就觉得有些内容很...

在浏览 scikit-learn 时 无意发现之，贡献者之一 是清华的Wei Li <https://github.com/kuantkid/PRML> 发现了，分享了；未曾使用。

大家用用看，看看楼下怎么说：

这本书最近开源了：

[<https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/>]

作为上课的教材读的，内容结构上比较全面。从基本的问题出发，对于每一个问题和范式的来由解释得比较详细清楚，也因而显得小章节间的逻辑关系 (有时) 堆得比...

在研一的下学期的时候，看了前三章。写得非常好，看着就不想放下。后来由于有其他事，就先停了下来。现在经过一年的实习，对机器学习感觉也算入门了，准备着手再开始看，相信这次会有完全不同的感觉。大家一起加油，PRML真是经典！

我们已经读完了Pattern Recognition And Machine Learning

，写的非常优美的一本书，另外我们正准备读MLAPP，欢迎加群177217565讨论。请在群申请理由里用简短的话描述一个算法的关键思想。

从大四就想看这本书，一直当做宝贝供着。。。。最近才大概翻了一遍，总体评价。。神书无疑。。读一遍感觉不行，我肯定还要读第二遍，因为有些章节还是有难度的。。作者写作功底太好，每个公式解释的都很清楚，看起来毫不费力，也很全面。总之，吐血推荐，不看这本书，别跟我说...

听完coursera上的机器学习的课后觉得ML不过就是拟合函数，但是由于里面并没有详细介绍算法背后的数学推理和我以为会有的关于“学习”“智能”的理论，就找来这本评价较好的书看了一遍。看完这本书后发现ML真的就是拟合函数，而且用的都是一些百年前的数学技巧和几十年前的算法...

Pattern Recognition and Machine Learning第九章读书会记录，主要内容：k-means 混合高斯 EM算法 <http://weibo.com/1841149974/A1O6GzO3k?mod=weibotime>

赞扬已经够多了，引用黄亮的话来说下这本书不好的地方。“这书把machine learning搞得太复杂太琐碎了，而迷失了其数学真意。其数学真意应该是简单统一的几何意义，而不是满屏的公式。另外这书理论深度不够，很多重要但简单的证明没讲。简言之，这书是电子工程师写的，不是给...

PRML读书会一周年资源汇总：http://weibo.com/p/10080817a99a8dcd9c7e83da56c7e13ede62a/emceercd?from=page_huati_rcd_more

[Pattern Recognition and Machine Learning_下载链接1](#)