

# 统计学习基础



[统计学习基础\\_下载链接1](#)

著者:哈斯蒂 (Hastie.T.)

出版者:世界图书出版公司

出版时间:2009-1-1

装帧:平装

isbn:9787506292313

计算和信息技术的飞速发展带来了医学、生物学、财经和营销等诸多领域的海量数据。

理解这些数据是一种挑战，这导致了统计学领域新工具的发展，并延伸到诸如数据挖掘、机器学习和生物信息学等新领域。许多工具都具有共同的基础，但常常用不同的术语来表达。《统计学习基础(第2版)(英文)》介绍了这些领域的一些重要概念。尽管应用的是统计学方法，但强调的是概念，而不是数学。许多例子附以彩图。《统计学习基础(第2版)(英文)》内容广泛，从有指导的学习（预测）到无指导的学习，应有尽有。包括神经网络、支持向量机、分类树和提升等主题，是同类书籍中介绍得最全面的。

《统计学习基础(第2版)(英文)》可作为高等院校相关专业本科生和研究生的教材，对于统计学相关人员、科学界和业界关注数据挖掘的人，《统计学习基础(第2版)(英文)》值得一读。

作者介绍:

作者：（德国）T.黑斯蒂（Trevor Hastie）

目录: Preface

1 Introduction Overview of Supervised Learning

2.1 Introduction

2.2 Variable Types and Terminology

2.3 Two Simple Approaches to Prediction: Least Squares and Nearest Neighbors

2.3.1 Linear Models and Least Squares

2.3.2 Nearest-Neighbor Methods

2.3.3 From Least Squares to Nearest Neighbors

2.4 Statistical Decision Theory

2.5 Local Methods in High Dimensions

2.6 Statistical Models, Supervised Learning and Function Approximation

2.6.1 A Statistical Model for the Joint Distribution  $\Pr(X, Y)$

2.6.2 Supervised Learning

2.6.3 Function Approximation

2.7 Structured Regression Models

2.7.1 Difficulty of the Problem

2.8 Classes of Restricted Estimators

2.8.1 Roughness Penalty and Bayesian Methods

2.8.2 Kernel Methods and Local Regression

2.8.3 Basis Functions and Dictionary Methods

2.9 Model Selection and the Bias-Variance Tradeoff

Bibliographic Notes

Exercises

3 Linear Methods for Regression

3.1 Introduction

3.2 Linear Regression Models and Least Squares

3.2.1 Example: Prostate Cancer

3.2.2 The Gauss-Markov Theorem

3.3 Multiple Regression from Simple Univariate Regression

3.3.1 Multiple Outputs

3.4 Subset Selection and Coefficient Shrinkage

3.4.1 Subset Selection

3.4.2 Prostate Cancer Data Example (Continued)

3.4.3 Shrinkage Methods

3.4.4 Methods Using Derived Input Directions

3.4.5 Discussion: A Comparison of the Selection and Shrinkage Methods

3.4.6 Multiple Outcome Shrinkage and Selection

### 3.5 Computational Considerations

Bibliographic Notes

Exercises

## 4 Linear Methods for Classification

### 4.1 Introduction

### 4.2 Linear Regression of an Indicator Matrix

### 4.3 Linear Discriminant Analysis

#### 4.3.1 Regularized Discriminant Analysis

#### 4.3.2 Computations for LDA

#### 4.3.3 Reduced-Rank Linear Discriminant Analysis

### 4.4 Logistic Regression

#### 4.4.1 Fitting Logistic Regression Models

#### 4.4.2 Example: South African Heart Disease

#### 4.4.3 Quadratic Approximations and Inference

#### 4.4.4 Logistic Regression or LDA?

### 4.5 Separating Hyper planes

#### 4.5.1 Rosenblatts Perceptron Learning Algorithm

#### 4.5.2 Optimal Separating Hyper planes

Bibliographic Notes

Exercises

## 5 Basis Expansions and Regularization

### 5.1 Introduction

### 5.2 Piecewise Polynomials and Splines

#### 5.2.1 Natural Cubic Splines

#### 5.2.2 Example: South African Heart Disease (Continued)

#### 5.2.3 Example: Phoneme Recognition

### 5.3 Filtering and Feature Extraction

### 5.4 Smoothing Splines

#### 5.4.1 Degrees of Freedom and Smoother Matrices

### 5.5 Automatic Selection of the Smoothing Parameters

#### 5.5.1 Fixing the Degrees of Freedom

#### 5.5.2 The Bias-Variance Tradeoff

### 5.6 Nonparametric Logistic Regression

### 5.7 Multidimensional Splines

### 5.8 Regularization and Reproducing Kernel Hilbert Spaces . .

#### 5.8.1 Spaces of Functions Generated by Kernels

#### 5.8.2 Examples of RKHS

### 5.9 Wavelet Smoothing

#### 5.9.1 Wavelet Bases and the Wavelet Transform

#### 5.9.2 Adaptive Wavelet Filtering

Bibliographic Notes

Exercises

Appendix: Computational Considerations for Splines

Appendix: B-splines

Appendix: Computations for Smoothing Splines

## 6 Kernel Methods

### 6.1 One-Dimensional Kernel Smoothers

#### 6.1.1 Local Linear Regression

#### 6.1.2 Local Polynomial Regression

### 6.2 Selecting the Width of the Kernel

### 6.3 Local Regression in Japan

### 6.4 Structured Local Regression Models in Japan

#### 6.4.1 Structured Kernels

#### 6.4.2 Structured Regression Functions

- 6.5 Local Likelihood and Other Models
- 6.6 Kernel Density Estimation and Classification
  - 6.6.1 Kernel Density Estimation
  - 6.6.2 Kernel Density Classification
  - 6.6.3 The Naive Bayes Classifier
- 6.7 Radial Basis Functions and Kernels
- 6.8 Mixture Models for Density Estimation and Classification
- 6.9 Computational Considerations
- Bibliographic Notes
- Exercises
- 7 Model Assessment and Selection
  - 7.1 Introduction
  - 7.2 Bias, Variance and Model Complexity
  - 7.3 The Bias-Variance Decomposition
    - 7.3.1 Example: Bias-Variance Tradeoff
  - 7.4 Optimism of the Training Error Rate
  - 7.5 Estimates of In-Sample Prediction Error
  - 7.6 The Effective Number of Parameters
  - 7.7 The Bayesian Approach and BIC
  - 7.8 Minimum Description Length
  - 7.9 Vapnik Chernovenkis Dimension
    - 7.9.1 Example (Continued)
  - 7.10 Cross-Validation
  - 7.11 Bootstrap Methods
    - 7.11.1 Example (Continued)
- Bibliographic Notes
- Exercises
- 8 Model Inference and Averaging
  - 8.1 Introduction
  - 8.2 The Bootstrap and Maximum Likelihood Methods
    - 8.2.1 A Smoothing Example
    - 8.2.2 Maximum Likelihood Inference
    - 8.2.3 Bootstrap versus Maximum Likelihood
  - 8.3 Bayesian Methods
  - 8.4 Relationship Between the Bootstrap and Bayesian Inference
  - 8.5 The EM Algorithm
    - 8.5.1 Two-Component Mixture Model
    - 8.5.2 The EM Algorithm in General
    - 8.5.3 EM as a Maximization-Maximization Procedure
  - 8.6 MCMC for Sampling from the Posterior
  - 8.7 Bagging
    - 8.7.1 Example: Trees with Simulated Data
  - 8.8 Model Averaging and Stacking
  - 8.9 Stochastic Search: Bumping
- Bibliographic Notes
- Exercises
- 9 Additive Models, Trees, and Related Methods
  - 9.1 Generalized Additive Models
    - 9.1.1 Fitting Additive Models
    - 9.1.2 Example: Additive Logistic Regression
    - 9.1.3 Summary
  - 9.2 Tree Based Methods
- 10 Boosting and Additive Trees
- 11 Neural Networks

12 Support Vector Machines and Flexible Discriminants  
13 Prototype Methods and Nearest-Neighbors  
14 Unsupervised Learning  
References  
Author Index  
Index  
• • • • • ([收起](#))

[统计学习基础\\_下载链接1](#)

## 标签

机器学习

统计学习

统计学

数据挖掘

数学

统计

概率论与统计学

Statistical

## 评论

课本 自身功力太浅

-----  
非常好的一本书，读了好几遍，每次都有新收获。只可惜我现在已经决定退坑了。

-----  
Will be a classic

-----  
Statistical Learning 最经典的入门教材。很多Machine Learning的书，没有关注太多模型背后的原理。但是要做好Machine Learning，这些必不可少。另，Youtube上可以搜到Hastie和小伙伴前些年的授课视频。

-----  
[统计学习基础\\_下载链接1](#)

## 书评

个人觉得“机器学习 -- 从入门到精通”可以作为这本书的副标题。机器学习、数据挖掘或者模式识别领域有几本非常流行的教材，比如Duda的模式分类，Bishop的PRML。Duda的书第一版是模式识别的奠基之作，现在大家谈论得是第二版，因为内容相对简单，非常流行，但对近20年取得统...

-----  
评论最下面的部分Version 1是我开始读这本书的时候写的东西，现在加上点基础部分。对linear algebra, probability要有非常强的直观认识，对这两个基础学的非常通透。Linear algebra有几种常用的分解QR, eigendecomposition, SVD，搞清楚它们的作用和几何意义。Bayesian meth...

-----  
有人给我推荐这本书的时候说，有了这本书，就不再需要其他的机器学习教材了。入手这本书的接下来两个月，我与教材中艰深的统计推断、矩阵、数值算法、凸优化等数学知识展开艰苦的斗争。于是我明白了何谓“不需要其他的机器学习教材”：准确地说，是其他的教材都不需要了；一本...

-----  
我导师(stanford博士毕业)非常欣赏这本书，并把它作为我博士资格考试的参考教材之一。感谢 ZHENHUI LI提供的信息。本书作者已经将第二版的电子书放到网上，大家可以免费下载。  
<http://www-stat.stanford.edu/~tibs/ElemStatLearn/> 网上还有一份solution manual,

但是似乎...

-----  
-----  
-----  
douban评论非要给出评价才能发表，这非常难决断 说你好呢，翻译的乱七八糟  
说你不好呢，内容实在深刻 说起翻译来，这可是把中文说的比外文还难懂 Jiawei  
Han的数据挖掘让范明译的污七八糟  
结果还让他来翻译这部经典，怀疑他在用google翻译  
最后还是忍不住去图书馆复印了原版...

-----  
这个简单的书评只是我个人的观点，所以我觉得先了解一下我的背景是有帮助的：本科  
计算机，数学功底尚可，研究生方向机器学习、数据挖掘相关应用研究。缺点：  
1，阅读此书前，读者需要具备基本的统计学知识，所以书的内容并不“基础”。  
2，书中很少涉及到公式推导，细节并不...

-----  
<https://esl.hohoweiya.xyz/index.html>  
=====

-----  
中文翻译版大概是用google翻译翻的，然后排版一下，就出版了。所以中文翻译版中，  
每个单词翻译是对的，但一句话连起来却怎么也看不懂。最佳阅读方式是，看英文版，  
个别单词不认识的话，再看中文版对应的那个词。但如果英文版整个句子都不懂的话，  
那只有去借助baidu/google，并...

-----  
读 ESL  
快半年了，也读了差不多1/3，写个短评记录一下，等读完的时候再来改吧。然后简单  
对比下基本常见的机器学习教材。  
我本科是学物理的，对于统计甚至概率论可以说是一无所知。入门的时候读的是周志华  
老师的《机器学习》，不过并没有读完的。一方面在家看书效率太低；另一...

-----  
[http://www-stat.stanford.edu/~hastie/local.ftp/Springer/ESLII\\_print3.pdf](http://www-stat.stanford.edu/~hastie/local.ftp/Springer/ESLII_print3.pdf)

-----  
对于新手来说，这本书和PRML比起来差太远，新手强烈建议去读PRML，接下来再看这本书。。我就举个最简单的例子吧，这本书的第二章overview of supervised learning和PRML的introduction差太远了。。。读这本书的overview如果读者没有基础几乎不知所云。。但是PRML通过一个例子...

-----  
非常难，一点都不element，是本百科全书式的读物，如果是初学者，不建议读很多章节也没有细节，概述性的东西，能看懂几章就很不错了  
其实每章都可以写成一本书，都可以做很多篇的论文  
全部读懂非常非常难，倒是作为用到哪个部分作为参考资料查查很不错

-----  
<https://web.stanford.edu/~hastie/ElemStatLearn/>  
=====

-----  
读了一个月，还在前四章深耕，在此说明一下，网上的solution，笔记啊，我见到的，只有一个份做的最详细，准确度最高，其余的都是滥竽充数，过程推导乱来，想当然，因为该书的符号有点混乱，所以建议阅读该书的人把前面的Notation读清楚，比如书中X出现的有好几种形式，每...

-----  
英文原版的官方免费下载链接已经有人在书评中给出了  
中文版的译者很可能没有基本的数学知识，而是用Google翻译完成了这部作品。  
超平面的Normal equation  
(法线方程) 翻译成了“平面上的标准方程”；而稍有高中高维几何常识的人都知道，法线是正交与该超平面的方向，而绝不可...

-----  
上半部看得更仔细些，相对来说收获也更多。书的前半部对各种回归说得很多，曾经仅仅了解这些的回归方法的大概思路，但是从本书中更能了解它们的统计意义、本质，有种豁然开朗的感觉：)  
只是总的来说还是磕磕巴巴的看了一遍，还得继续仔细研读才好。希望能有更深刻的领悟，目的...

-----  
统计学习的经典教材，数学难度适中，英文难度较低，看了其中有监督学习部分，无监督学习部分没怎么看，算法比较经典，但是也比较老。

-----  
The methodology used in the books are fancy and attractive, yet in terms of rigorous proofs, sometimes the book skip steps and is difficult to follow. ~ Slightly sophisticated for undergraduate students, but in general is a very nice book.

-----  
[统计学习基础\\_下载链接1](#)