

# 参考教材



## 数据挖掘算法原理与实现

王振武，徐慧 著  
清华大学出版社，2015年2月

# 参考教材



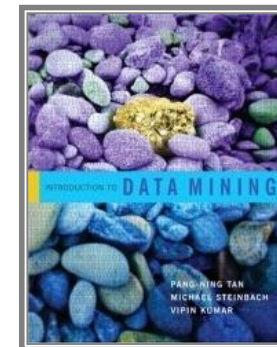
## 数据挖掘概念与技术（原书第2版）

韩家炜，堪博 著，范明，孟小峰 译  
机械工业出版社，2007年3月  
ISBN: 9787111205388



## Introduction to Data Mining

by [Pang-Ning Tan](#), Michael Steinbach, Vipin Kumar  
Addison Wesley; May 2005  
ISBN-13: 978-0321321367



# 推荐资源



- 全球最大的数据挖掘信息网站——  
<http://www.kdnuggets.com>
  - 数据挖掘和分析软件，求职，咨询，课程等数据挖掘社区资源
- 机器学习领域的UCI数据集——  
<http://archive.ics.uci.edu/ml>
  - University of California Irvine 提出的用于机器学习的数据库，目前拥有200多个数据集，并且数目还在不断增加。
  - 在数据挖掘领域被认为是标准测试数据集。



# 推荐资源

---

- UCI数据集: <http://kdd.ics.uci.edu/>
- CMU数据集: <http://lib.stat.cmu.edu/datasets/>  
<http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/>
- 时序数据集: <http://www.stat.wisc.edu/~reinsel/bjr-data/>
- 金融数据集: <http://lisp.vse.cz/pkdd99/Challenge/chall.htm>
- 癌症基因数据集: <http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi>
- 综合数据集: <http://www.cs.nyu.edu/~roweis/data.html>
- 数据集列表: <http://www.kdnuggets.com/datasets/index.html>



# 推荐资源

---

- UCI机器学习网站 <http://archive.ics.uci.edu/ml/>
- Weka官方网站 <http://www.cs.waikato.ac.nz/ml/weka/>
- DBMiner官方网站 <http://ddm.cs.sfu.ca/>
- SVM代码 <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- 其它开源软件包： NB（朴素贝叶斯网络）， NN（神经网络）， DT（决策树）
- 相关软件： Matlab， StatSoft等商用软件； SQL Server 2008中也提供了相应的Data Analysis数据分析工具

# 学时分配

表 0-1 学时安排

教学内容	学时分配			小计
	讲课	上机实践	讨论/习题	
第0讲 入课	1		1	2
第1讲 认识数据挖掘	5		2	7
第2讲 基本数据挖掘技术	6	3		9
第3讲 数据库中的知识发现	3	3		6
第4讲 数据仓库	3	3	3	9
第5讲 评估技术	3		3	6
第6讲 神经网络技术	3	3		6
第7讲 统计技术	3	3	3	9
第8讲 时间序列和基于Web的挖掘	3		3	6
合 计	30	15	15	60