



Mathematics in computer science

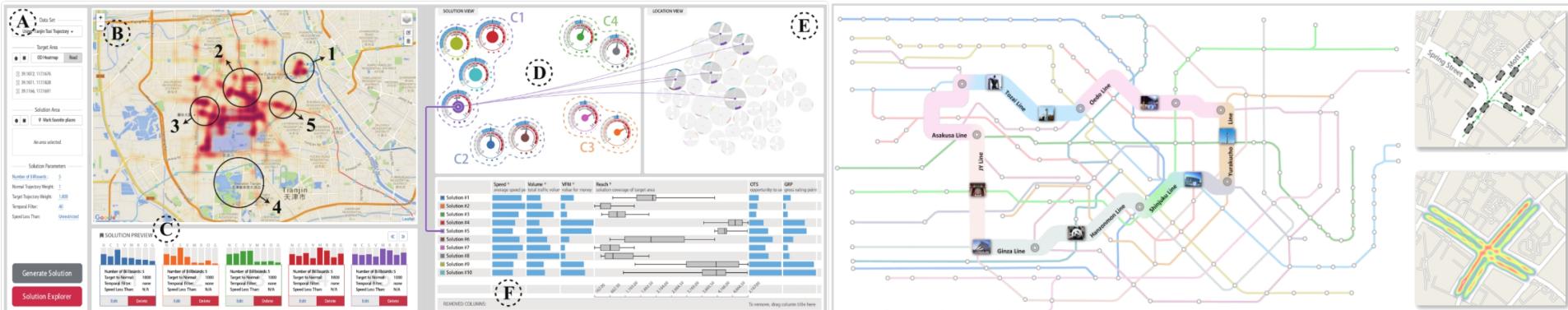
Introduction

张宏鑫 (Hongxin Zhang)

zhx@cad.zju.edu.cn

State Key Lab of CAD&CG, ZJU

2025-02-18





Syllabus

- This 2 months short course:
 - covers a set of data driven techniques
 - optimization methods
 - from basic and state-of-the-art.
- You will learn fundamental algorithms of multivariate analysis
- And see the stories behind these algorithms, theory and applications.
- It is going to be fun and hard work.



Rough schedule

- 02.18: Introduction & Point estimation
 - 02.25: Component Analysis
 - 03.04: Distance and similarity
 - 03.11: Graphical models
-
- 03.18: Linear programming (I)
 - 03.25: Linear programming (II)
 - 04.01: Non-linear programming
 - 04.08: Quadratic programming



16岁高中生的“卷”：用13000+行代码，从头写了一个C++机器学习库

开源前哨 2022-02-20 11:40

来源：机器之心

人工智能领域现在也流行高中生拯救世界了？

一个热爱计算机的少年，16岁就已经可以做出点东西来了，比如开发个粤语编程语言、拿个Kaggle冠军、写个游戏、开发个加密货币投资机器人、从头构建一个C++机器学习库什么的。

今天要介绍的就是一位从头构建C++机器学习库的16岁少年(@novak-99)，他的自荐帖在reddit上获得了数百的点赞量。

A screenshot of a reddit post from the user u/novak-99. The post is titled "[P] C++ Machine Learning Library Built From Scratch by a 16-Year-Old High Schooler". It has 357 upvotes and is categorized under "Project". The post begins with "Hello r/MachineLearning!", followed by a statement explaining the creation of a machine learning library in C++ from scratch. It concludes with a link to the GitHub repository and a note about starting work at age 15.

Posted by u/novak-99 20 hours ago

357 [P] C++ Machine Learning Library Built From Scratch by a 16-Year-Old High Schooler

Project

Hello [r/MachineLearning!](#)

In this post, I will be explaining why I decided to create a machine learning library in C++ from scratch.

If you are interested in taking a closer look at it, the GitHub repository is available here: <https://github.com/novak-99/MLPP>. To give some background, the library is over 13.0K lines of code and incorporates topics from statistics, linear algebra, numerical analysis, and of course, machine learning and deep learning. I have started working on the library since I was 15.

他构建的这个库（ML++）有13000多行代码，涵盖了统计、线性代数、数值分析、机器学习和深度学习等主题。



目前，ML++ 库中正在开发以下模型和技术：

- 卷积神经网络（CNN）
- 支持向量机（SVM）的内核
- 支持向量回归

<https://github.com/novak-99/MLPP>

整体而言，ML++ 库包含了 19 大主题以及相关细分内容，分别如下：

- 回归（线性回归、逻辑回归、Softmax 回归、指数回归、Probit 回归、Cloglog 回归、Tanh 回归）
- 深度、动态、规模化神经网络（激活函数、优化算法、损失函数、正则化方法、权重初始化方法、学习率规划器）
- Prebuilt 神经网络（多层感知机、自编码器、Softmax 网络）
- 生成建模（表格对抗生成网络）
- 自然语言处理（Word2Vec、词干提取、词袋模型、TFIDF、辅助文本处理函数）
- 计算机视觉（卷积操作、最大 / 最小 / 平均池化、全局最大 / 最小 / 平均池化、Prebuilt 特征向量）
- 主成分分析
- 朴素贝叶斯分类器（多项分布朴素贝叶斯、伯努利分布朴素贝叶斯、高斯分布朴素贝叶斯）
- 支持向量分类（原始形成、对偶形成）
- K-Means 算法
- K 最近邻算法
- Outlier Finder（使用标准分数）
- 矩阵分解（SVD 分解、Cholesky 分解、QR 分解）
- 数值分析（数值微分、Jacobi 向量计算器、Hessian 矩阵计算器、函数近似器、微分方程求解器）
- 数学变换（离散余弦变换）
- 线性代数模块
- 统计模块
- 数据处理模块（特征缩放、均值归一化、One Hot 表征、反 One Hot 表征、支持的颜色空间转换类型）
- 实用工具（TP/FP/TN/FN 函数、精度、召回率、准确率、F1 分数）



Self-reflection 反思

- It is NOT a machine learning course
 - Although we will discuss a lot on machine learning things
- We will focus on mathematical methods and their underlying motivation
 - Representation and presentation
 - Thinking in mathematical way
 - Happy working with mathematics

Principle

- Simple is beauty!
- Make a balance between theories and real applications
- 大道无形：大音希声，大象无形，道隐无名。



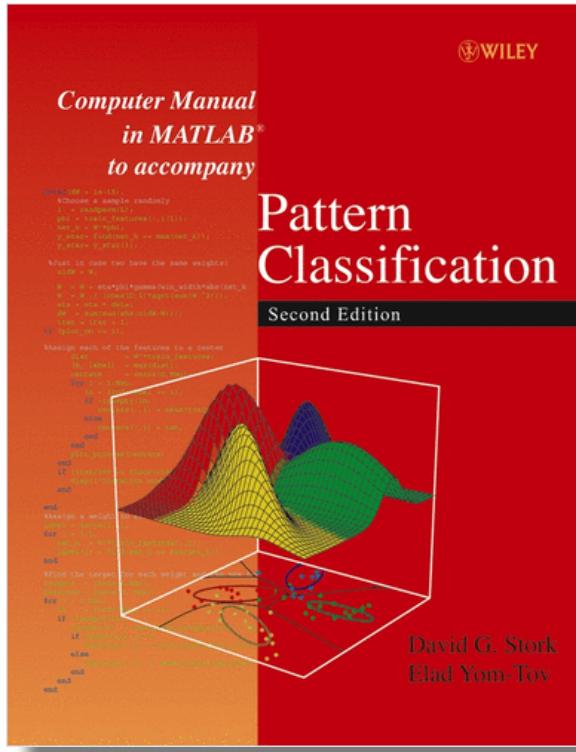


Prerequisites

- Probabilities
 - Distributions, densities, marginalization...
- Basic statistics
 - Moments, typical distributions, regression...
- Algorithms
 - Dynamic programming, basic data structures, complexity...
- Programming
 - Mostly your choice of language: C/C++, MATLAB, JAVA
- We provide some background, but the class will be fast paced
- Ability to deal with “abstract mathematical concepts”

Text books

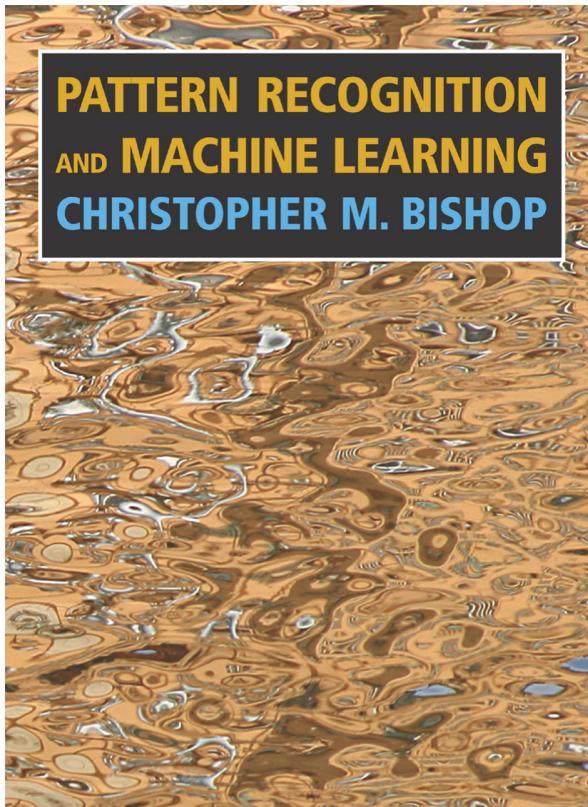
- Pattern Classification (2nd Edition)
 - by Duda, Hart and Stork
- Information Theory, Inference, and Learning Algorithms
 - by David MacKay





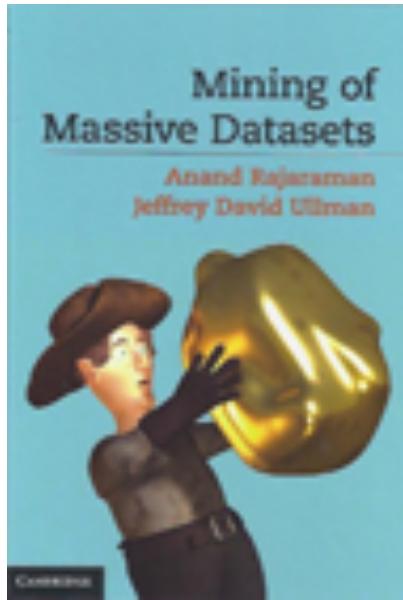
Text books

- <http://research.microsoft.com/en-us/um/people/cmbishop/prml/>



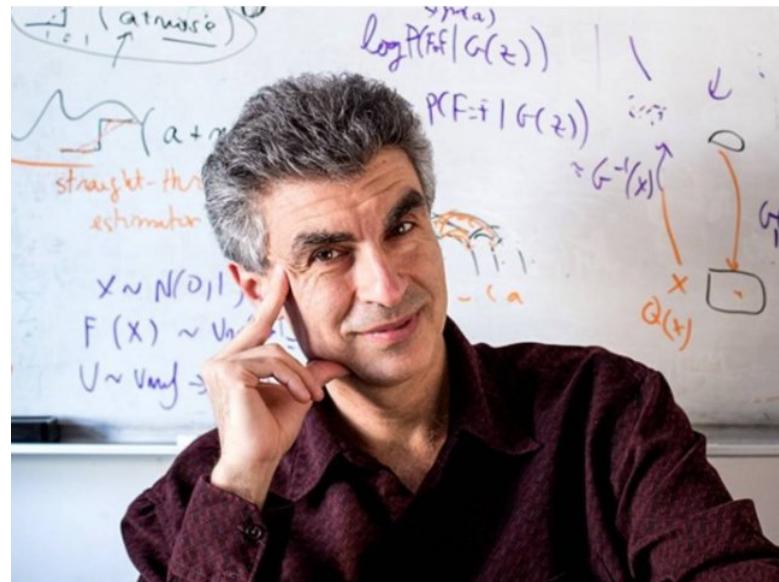
Text books

- From data mining area
 - Mining of Massive Datasets
 - <http://infolab.stanford.edu/~ullman/mmds.html>



Text books

- From the HOT deep learning field:
- Written by Yoshua Bengio
 - <http://www.deeplearningbook.org/>
 - <https://github.com/exacity/deeplearningbook-chinese>





- <https://github.com/hongxin/csmath-2025>

Reference

[https://github.com/ChristosChristofidis/
awesome-deep-learning](https://github.com/ChristosChristofidis/awesome-deep-learning)



Evaluation

- Homework: 40%
 - Python programming (30%)
 - Course notes
 - In class performance: (10%)
 - check-in
 - Short talk (optional)
- Course paper: 60%
 - Read top-level (10 selected) papers
 - and report main idea



Homework 0: registration

- Say hello to me and TA:
 - Link: course.zju.edu.cn
 - Title:
 - Content: say something about yourself
 - Attachment: one CLEAR frontal photo of yourself, we will use it for further functions based on face-recognition



Alternative ~

14:43 5G

< ⌂ ...



计算机应用数学课助教

数学助教，博士研究生，擅长大数据可视化，辅导学生学习。

1人聊过 · 创建人 @Tim

分享智能体

查找聊天内容 >

智能体设定 >

声音 沉稳皓轩 >

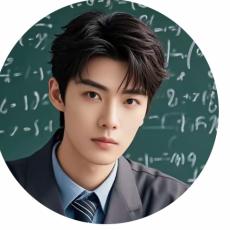
语言 中文 >

清除上下文

由豆包大模型提供支持

14:49 5G

< 计算机应用数学课... @Tim ⌂ ...



我是数学助教，可答疑、协助写程序和论文。有问题随时问，比如“这个算法怎么理解”。

请谈谈贝叶斯统计

贝叶斯统计是一种基于贝叶斯定理的统计方法。

贝叶斯定理的核心是通过先验概率和条件概率来计算后验概率。

先验概率是在没有新的证据或信息之前，对某个事件发生概率的 ↓ 估计。而条件概率

发消息... ⌂



源自88上某位网友的签名档

- 鉴于大多数博士们在之后的生活中并没有从事博士生期间的课题的研究，甚至根本不再做研究工作，我想攻读博士的目标应该是：
 1. 成为一个身体强壮的人
 2. 成为一个意志强悍的人
 3. 成为一个能系统思考，从混沌的一堆问题中提炼主要的具体的问题的人
 4. 成为一个能解决具体问题的人
- 修行！



- 业精于勤荒于嬉
- 行成于思毁于随

微博: @浙大张宏鑫

邮件: zhx@cad.zju.edu.cn

主页: <http://person.zju.edu.cn/zhx>

手机: 13958011790

微信: *timothykull*



谢谢

Thank You