作者华校专，曾任阿里巴巴资深算法工程师，现任智易科技首席算法研究员，《Python 大战机器学习》的作者。

这是作者多年以来学习总结的笔记，经整理之后开源于世。目前还有约一半的内容在陆续整理中，已经整理好的内容放置在此。 曾有出版社约稿，但是考虑到出版时间周期较长，而且书本购买成本高不利于技术广泛传播，因此作者采取开源的形式。 笔记内容仅供个人学习使用，非本人同意不得应用于商业领域。

笔记内容较多，可能有些总结的不到位的地方，欢迎大家探讨。联系方式:huaxz1986@163.com

另有个人在 github 上的一些内容：

- "《算法导论》的C++实现"代码：https://github.com/huaxz1986/cplusplus-_Implementation_Of_Introduction_to_Algorithms
- 《Unix 环境高级编程第三版》笔记：https://github.com/huaxz1986/APUE_notes

# 数学基础



# 统计学习

# 深度学习

# 自然语言处理



# 工具

## CRF

- CRF++
  - 一、安装
  - 二、使用
  - 三、Python接口
  - 四、常见错误

## lightgbm

- lightgbm使用指南
  - 一、安装
  - 二、调参
  - 三、进阶
  - 四、API
  - 五、Docker

## xgboost

- xgboost使用指南
  - 一、安装
  - 二、调参
  - 三、外存计算
  - 四、 GPU计算
  - 五、单调约束
  - 六、 DART booster
  - 七、Python API

## scikit-learn

- 1.预处理
  - 一、特征处理
  - 二、特征选择
  - 三、字典学习
  - 四、PipeLine
- 2.降维
  - 一、PCA
  - 二、MDS
  - 三、Isomap
  - 四、LocallyLinearEmbedding
  - 五、FA
  - 六、FastICA
  - 七、t-SNE
- 3.监督学习模型
  - 一、线性模型
  - 二、支持向量机
  - 三、贝叶斯模型
  - 四、决策树
  - 五、KNN

## spark

## numpy



## scipy



## matplotlib



## pandas