



Python

机器学习实战

课程介绍

机器学习发展的原动力

机器学习和数据分析的区别

算法分类 (1)

- 有监督学习
- 无监督学习
- 半监督学习

算法分类 (2)

- 分类和回归
- 聚类
- 关联分析

课程介绍

机器学习是个交叉学科

- 概率和统计
- 线性代数
- 微积分
- 算法
- 编程

机器学习解决问题的框架

- 定义目标
- 定义模型
- 定义损失函数
- 训练样本
- 优化

模型、目标和算法

课程介绍

课程计划

1、最近邻分类算法

2、贝叶斯分类算法

- 概率、贝叶斯定理

3、聚类算法

- 层次聚类法、Kmeans聚类法

4、决策树分类算法

- 决策树指标介绍——熵、信息增益、信息增益率

- ID3、ID4.5、ID5、CART四种决策

- 随机森林和装袋算法

5、线性回归

- 最小二乘法的含义

- 微积分和梯度下降算法

- 岭回归和lasso回归、弹性网

6、逻辑回归

- 广义线性回归和逻辑回归

- 逻辑回归中梯度的推导过程

课程介绍

课程计划

7、支持向量机

- 支持向量机的原理
- 三种支持向量机
- 核技术
- SMO算法

8、EM算法和GMM

- 什么是EM算法
- E步和M步到底要做些什么
- EM算法解决实际聚类问题

9、随机森林和Adaboost

- 什么是随机森林
- 可以和SVM相媲美的Adaboost
- Adaboost的精髓和实战

10、机器学习中的精华和技巧

- 各种机器学习算法的统一视角
- 过拟合问题和方差、偏差的关系
- 解决过拟合问题的正则化方法及L1、L2
- 实际问题中如何选择模型和参数
- 自适应学斜率算法

课程介绍

国际权威学术组织ICDM在
2006年12月评选出数据挖掘
领域的十大经典算法

数据挖掘十大算法：一览表



排名	挖掘主题	算法	得票数	发表时间	作者	讲解人
1	分类	C4.5	61	1993	Quinlan, J.R	Hiroshi Motoda
2	聚类	K-Means	60	1967	MacQueen, J.B	Joydeep Ghosh
3	统计学习	SVM	58	1995	Vapnik, V.N	Qiang Yang
4	关联分析	Apriori	52	1994	Rakesh Agrawal	Christos Faloutsos
5	统计学习	EM	48	2000	McLachlan, G	Joydeep Ghosh
6	链接挖掘	PageRank	46	1998	Brin, S.	Christos Faloutsos
7	集装与推进	AdaBoost	45	1997	Freund, Y.	Zhi-Hua Zhou
8	分类	kNN	45	1996	Hastie, T	Vipin Kumar
9	分类	Naïve Bayes	45	2001	Hand, D.J	Qiang Yang
10	分类	CART	34	1984	L.Breiman	Dan Steinberg