

机器学习导论

习题三

学号, 作者姓名, 邮箱

2017 年 4 月 12 日

1 [30pts] Decision Tree Analysis

决策树是一类常见的机器学习方法，但是在训练过程中会遇到一些问题。

(1) [10pts] 试证明对于不含冲突数据(即特征向量完全相同但标记不同)的训练集，必存在与训练集一致(即训练误差为0)的决策树；

(2) [10pts] 试分析使用“最小训练误差”作为决策树划分选择的缺陷。

Solution. 此处用于写证明(中英文均可)

2 [30pts] Training a Decision Tree

考虑下面的训练集：共计6个训练样本，每个训练样本有三个维度的特征属性和标记信息。详细信息如表1所示。

请通过训练集中的数据训练一棵决策树，要求通过“信息增益”(information gain)为准则来选择划分属性。请参考书中图4.4，给出详细的计算过程并画出最终的决策树。

表 1: 训练集信息

序号	特征 A	特征 B	特征 C	标记
1	0	1	1	0
2	1	1	1	0
3	0	0	0	0
4	1	1	0	1
5	0	1	0	1
6	1	0	1	1

Solution. 此处用于写解答(中英文均可)

3 [40pts] Back Propagation

单隐层前馈神经网络的误差逆传播(error BackPropagation, 简称BP)算法是实际工程实践中非常重要的基础, 也是理解神经网络的关键。

请编程实现BP算法, 算法流程如课本图5.8所示。详细编程题指南请参见链接: http://lamda.nju.edu.cn/ml2017/PS3/ML3_programming.html

在实现之后, 你对BP算法有什么新的认识吗? 请简要谈谈。

Solution. 此处用于写解答(中英文均可)

附加题 [30pts] Neural Network in Practice

在实际工程实现中, 通常会使用已有的开源库, 这样会减少搭建原有模块的时间。因此, 请使用现有神经网络库, 编程实现更复杂的神经网络。详细编程题指南请参见链接: http://lamda.nju.edu.cn/ml2017/PS3/ML3_programming.html

和上一题相比, 模型性能有变化吗? 如果有, 你认为可能是什么原因。同时, 在实践过程中你遇到了什么问题, 是如何解决的?

Solution. 此处用于写解答(中英文均可)