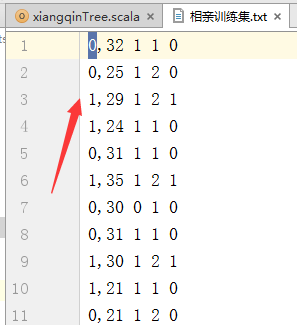
# 课堂笔记

易爱华 20:57:59

一定要信息增益比越大的特征作为根节点吗？

答：一定会选择信息增益最佳的作为根结点。

决策树案例：



第一列是Class，其他列是属性

如果Model的错误率较高，超过30%，如何优化？

答：1、检查训练集是否样本太少

1. 调整因子，观察对比。

郑博 21:43:39

會不會算法本身有問題需要改源代碼

答：不存在这个情况。

郑博 21:54:35

據説京東據説全部重寫

說是精度不夠

答：估计是忽悠。

易爱华 21:46:12

python数据挖掘相比于mllib数据挖掘的优缺点是什么？

答：phthon 不大清楚，深度学习库比较强大，如tenserFlow。

郑博 22:00:16

Tensorflow spark可以用不

* 答：我回头研究下，

雅虎开源

。

------------------

Mhount：基于MR计算引擎的，对常用数据挖掘算法进行了封装。