

对问题的回答

1. 你对 AdaBoost 算法有何新的认识？

从表中的数据来看，我对 AdaBoost 算法有了以下新的认识：

- 1) AdaBoost 作为分类器时，分类的精度比较高。两种经过训练的模型，最终的交叉验证准确率都在 90% 左右。
- 2) AdaBoost 的训练错误率上界，随着迭代次数的增加，会逐渐（呈指数）下降，这一点在决策树桩上可以明显看出，数据恰好呈一个上升的趋势。
- 3) AdaBoost 算法的训练时间比较长。算法实现过程中，从加权错误率向预期值逐渐逼近来构造级联分类器，迭代训练生成大量的弱分类器后才能实现这一构造过程，所以训练的时间比较长。
- 4) AdaBoost 算法不容易出现过拟合问题，但不是绝对的。当弱学习器的复杂度很大或训练数据含有较大的噪声时，可能会发生过拟合。

2. 关于基分类器类型、超参数设置对最终模型性能的影响，你有何发现？

关于基分类器类型、超参数设置对最终模型性能的影响，我发现：

- 1) 在 AdaBoost 的框架下，可以使用各种回归分类模型来构建弱学习器，最终得到的模型拟合效果都比较好，非常灵活。
- 2) 最终准确率和基分类器有关系。决策树准确率稳步上升，而对数几率回归准确率稳定，不符合上述认识，可能是因为基分类器的分类准确率已经很高了（本实验中对数几率回归不是一个弱分类器），后续产生了过拟合。
- 3) 模型的性能受所选超参数值的影响很大。对数几率模型中的主要超参数是学习率和迭代次数，但在该模型中迭代停止条件是梯度的模长足够小，学习率可以设定为固定值，不要过大或过小即可。AdaBoost 算法的主要超参数是基分类器数目，一般来讲，基分类器数目越大，准确率会越高。