**8.3 从网上下载或自己编程实现AdaBoost，以不剪枝决策树桩为基学习器在西瓜数据集3.0α上训练一个AdaBoost集成，并与图8.4进行比较。**

**①问题描述：**

本题要求使用AdaBoost算法以不剪枝决策树为基学习器利用西瓜数据集3.0α训练一个AdaBoost集成。但是若采用不剪枝决策树，则每个决策树分类器是趋于一致的，无法保证个体学习器的多样性。因此个体学习器应选用深度为2的单层决策树，即决策树桩作为弱学习器进行AdaBoost训练。

**②实现过程：**

a.将数据集中的密度及含糖率数据写死在代码中内的函数中转成numpy类型供后续调用。

b.开始对模型进行迭代。

c.建立决策树桩的基学习器并找到其中最佳的单层决策树。

d.将最佳的单层决策树加入到单层决策数组中。

e.计算权值α的值。

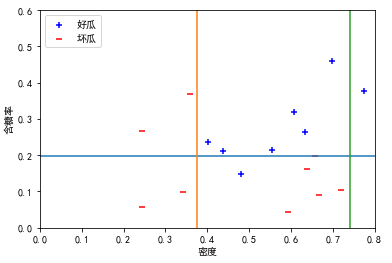
f.计算新的权重向量。

g.若错误率为0，则退出迭代循环，否则继续进行迭代，重复以上c、d、e、f步骤。

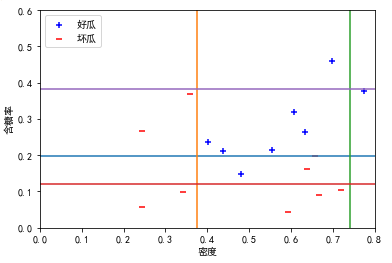
h.使用以上步骤中训练好的AdaBoost模型对测试样本进行预测，并绘制出图像。

**③实现结果：**

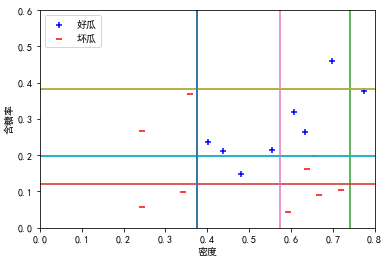
学习器数量=3，准确率= 0.9411764705882353



学习器数量=5，准确率= 0.9411764705882353



学习器数量=11，准确率= 1.0



**④结果分析：**

在本例的数据及学习器模型中，当基学习器数量为3和5时，其训练出的AdaBoost模型识别的准确率一致，说明基学习器数量增加的幅度偏小，不足以对模型性能进行提升。当基学习器数量提升至11时，其训练出的AdaBoost模型识别的准确率达到百分之百，说明基学习器数量的增加在一定程度上改善了模型的性能。与图8.4相比，可以看出，当基学习器数量较少时，本模型能够对样本进行更高准确率的识别，且当基学习器数量较多时，本模型对好瓜坏瓜的分类界限更为分明，在分类边界附近的样本数量较少，分割特征更加明显。