A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting *

Yoav Freund and Robert E. Schapire

■ 简介

Adaboost是一种基于boost思想的一种自适应的迭代式算法。 通过在同一个训练数据集上训练多个弱分类器(weak classifier)， 然后把这一组弱分类器组合起来， 产生一个强分类器（strong classifier）

■ 优点:

- 具有较低的泛化误差（low generalization）
- 改善了分类器的分类正确率
- 可以将这个算法配合多个分类算法使用。 例如， 可以选择决策树，SVM，贝叶斯分类器等作为弱分类器
- 容易编码实现，不容易出现overfiting(过拟合)现象

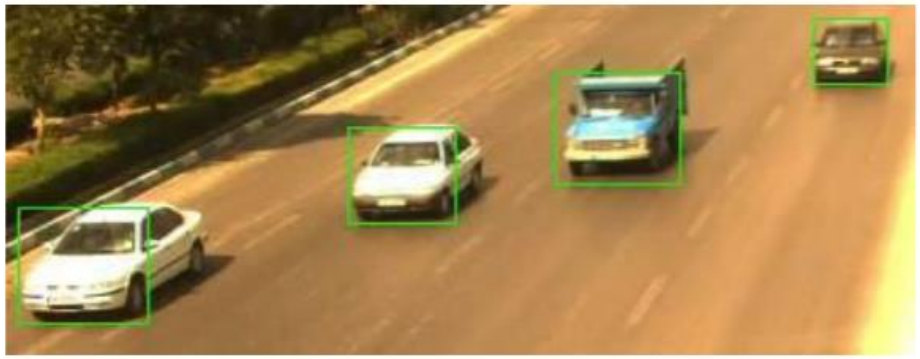
■ 缺点:

- 对outliers（异常点）非常的敏感。 因为异常点容易分错， 会逐级影响后面的产生的弱分类器

adaboost

■ 应用

medium lighting



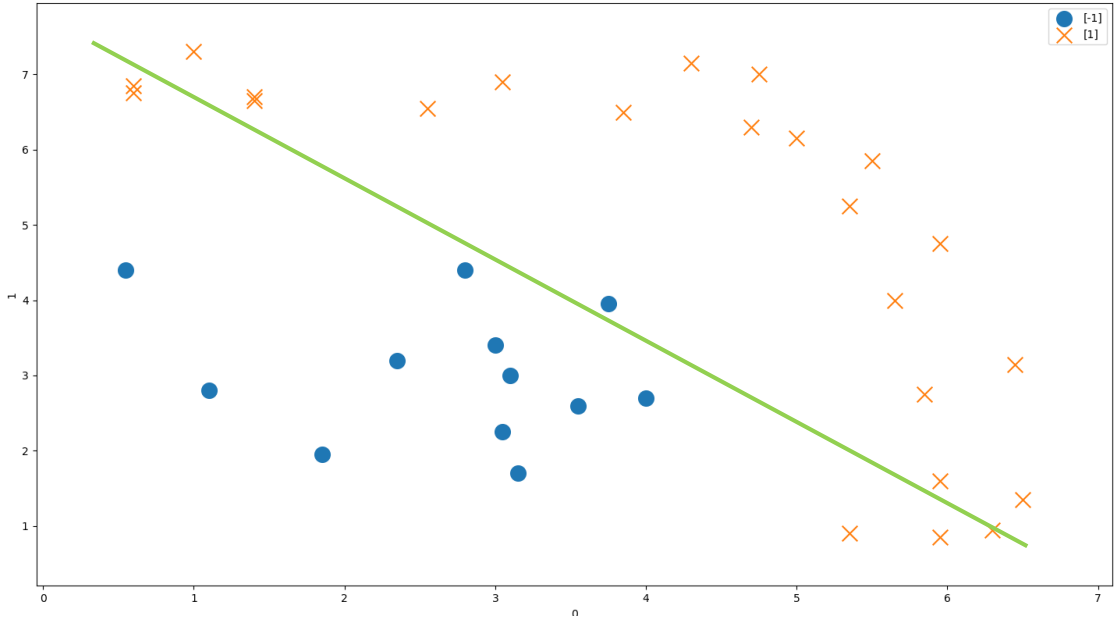
low lighting

strong lighting

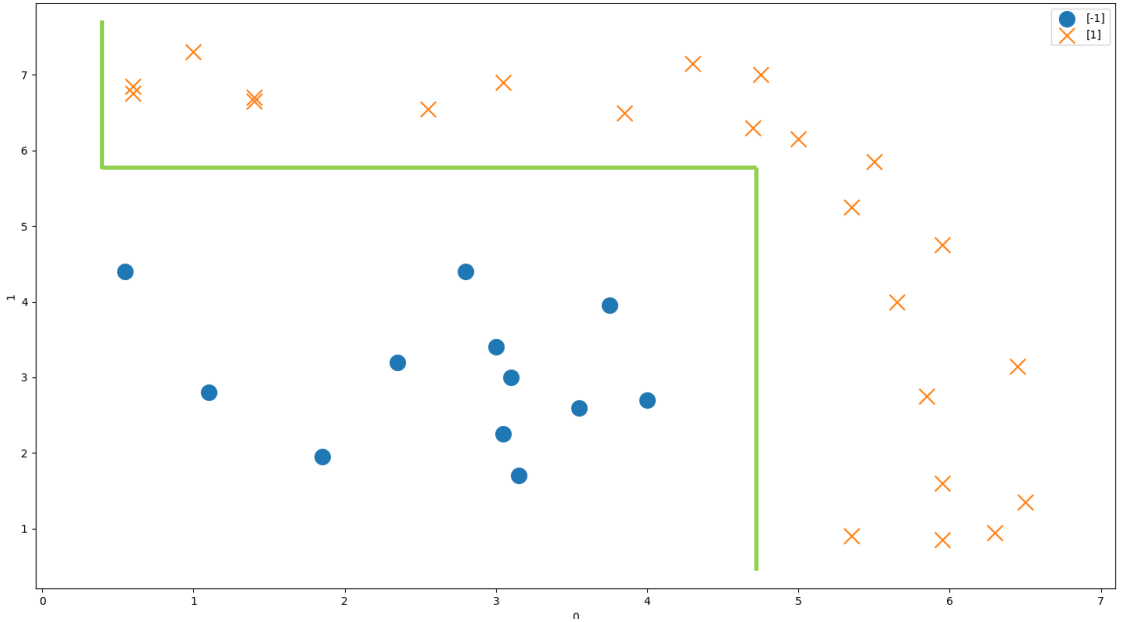
Mohammad, M., Maryam, N., Majid, P., Mohammad, K., 2018, Moving Vehicle Detection Using AdaBoost and Haar-Like Feature in Surveillance Videos, Islamic Azad University, Yazd, Iran

预备知识

■ 弱分类器

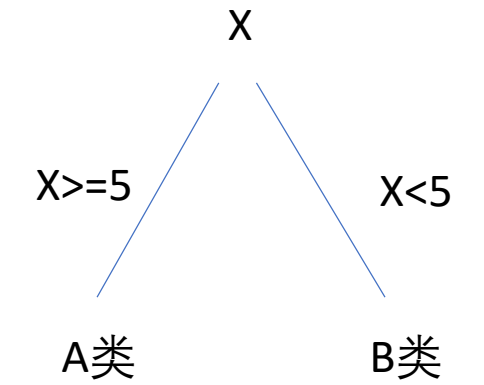


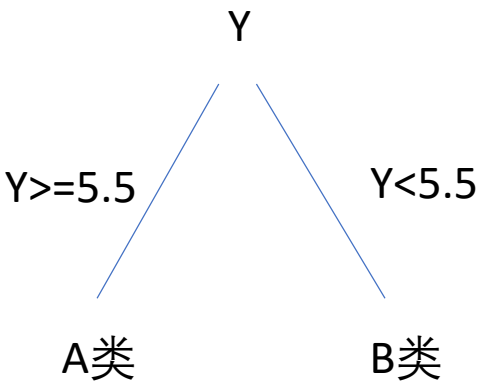
■ 强分类器

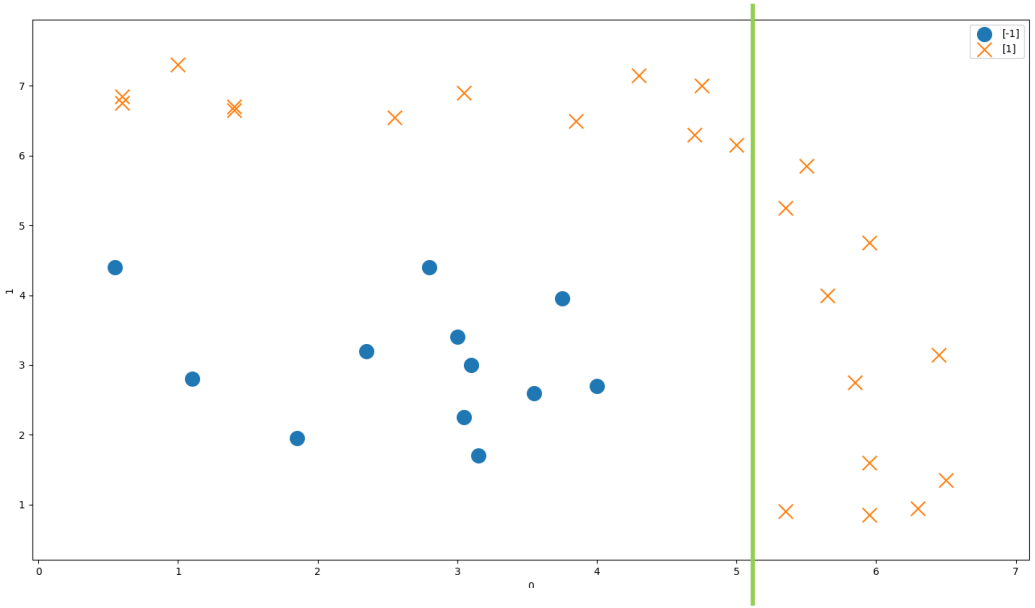


预备知识

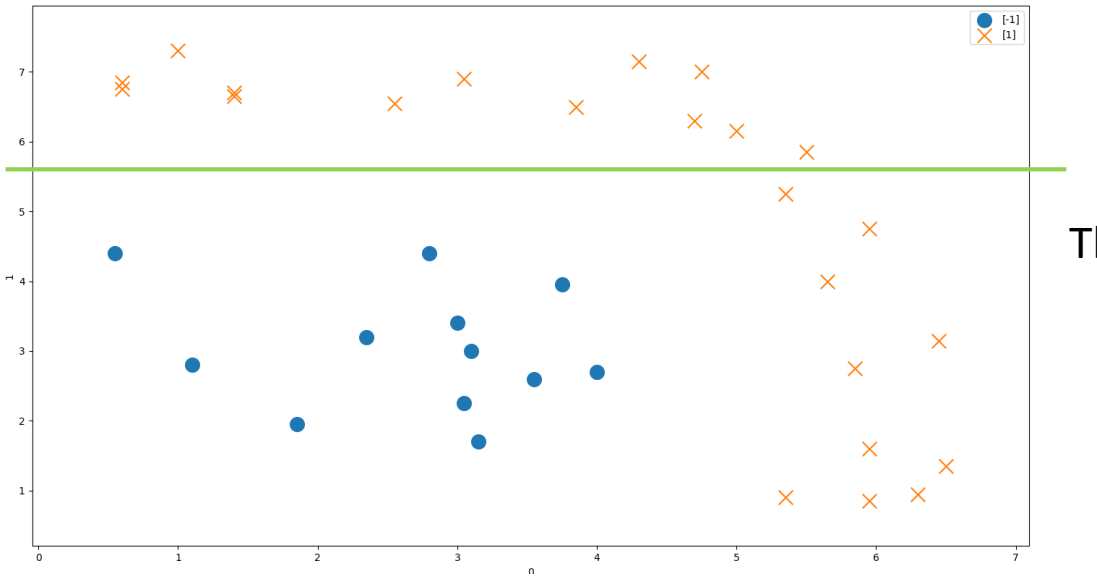
■ 单层决策树







Threshold x = 5



Threshold Y = 5.5

■ 算法基本思想

1）给所有训练样本增加一个权重属性w。

2）根据当前的样本，训练一个弱分类器（线性分类器或者决策树等）。用错分的样本的权重之和表示这个弱分类器的错误率e。

3）使用弱分类器的错误率更新样本的权重w。如果样本在该弱分类器中被正确分类，则减小其权重；若被错误分类，则增加其权重。或者只减小正确分类的样本的权重，这样相当于增加了错误分类样本的权重。

4）使用弱分类器的错误率生成弱分类器的权重α。

5）重复第二步，产生若干个弱分类器，直到所有分类器的加权α的分类的错误率e_f为0。

■ Adaboost算法分类

二分类	basic adaboost
多分类	adaboost M1 adaboost.M2
回归问题	adaboost.R

注：论文[1]给出了四种算法，并给出了一些基本的推导。论文[2] 给出了二分类算法的详细推导。

[1] A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*

[2] AdaBoost and the Super Bowl of Classifiers A Tutorial Introduction to Adaptive Boosting

注意：更新权重的两种方法是等价的。在归一化权重时，多的系数会被约掉，最后得到一样的权重。

◆ 二分类算法步骤

⓪ i表示第i个样本，共N个样本。M表示第 m次迭代，第一次迭代时，每个权重 W_1^i为$\frac{1}{N}$。且样本的标签为1和-1。$h_m(x_i)$表示第m个分类器对第i个样本的输出。

① 归一化权重$W_m^i = \frac{W_m^i}{\Sigma W_m^i}$

② 第m次迭代时构造第m个若分类器，使得

$$e_m = \sum_{y_i = h_m(x_i)} W_m^i$$

③ $\alpha_m = \frac{1}{2}\ln\frac{1-e_m}{e_m}$ 或者 $\alpha_m = \frac{e_m}{1-e_m}$

④ $W_{m+1}^i = W_m^i \cdot e^{-y_i \alpha_m h_m(x_i)} = \begin{cases} W_m^i \cdot \left(\frac{1-e_m}{e_m}\right)^{-\frac{1}{2}} \\ W_m^i \cdot \left(\frac{1-e_m}{e_m}\right)^{\frac{1}{2}} \end{cases}$

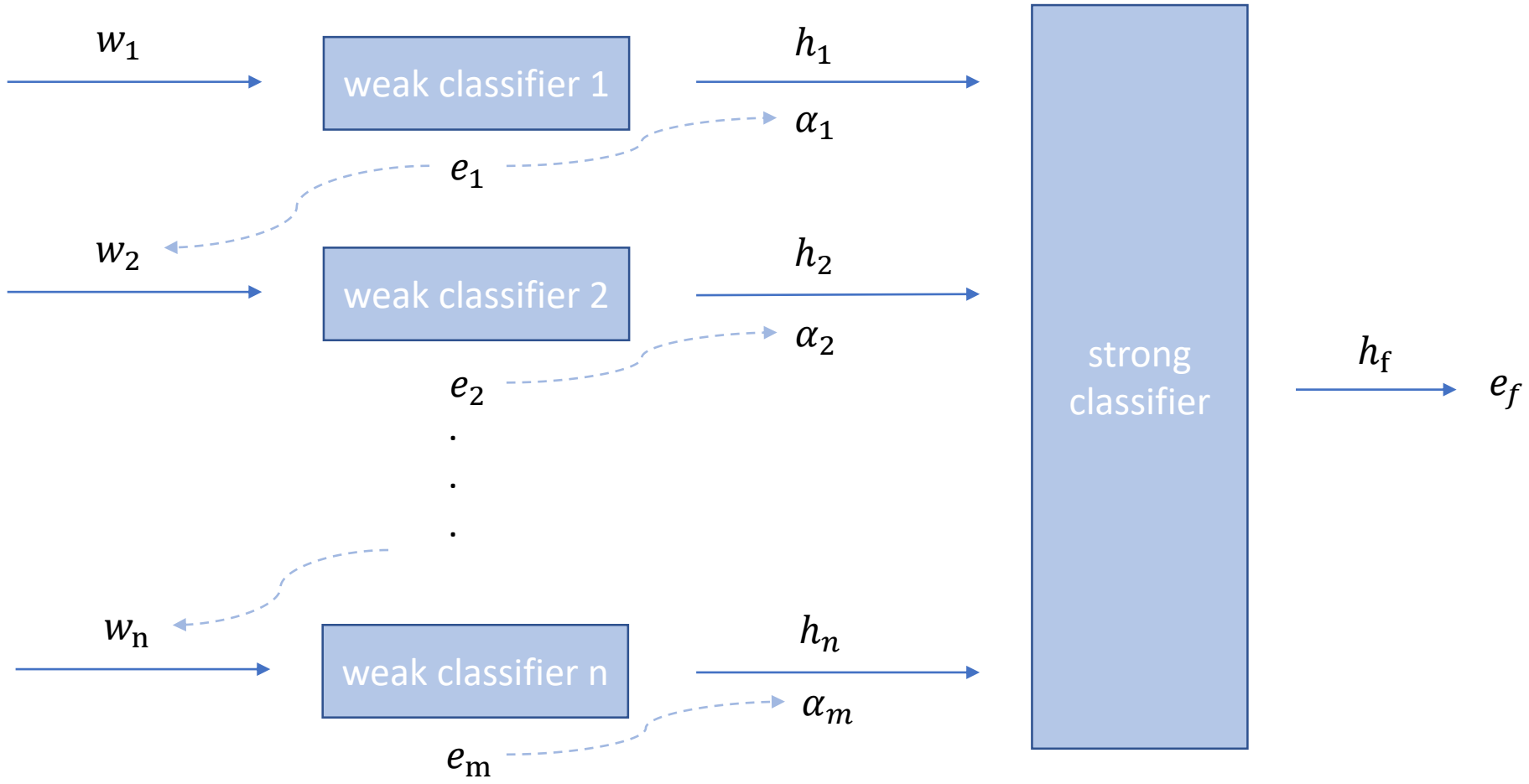
或者 $W_{m+1}^i = W_m^i \cdot \alpha_m{}^{1-|h_m(x_i)-y_i|/2}$

⑤ $h_f(x_i) = \mathrm{sgn}(\Sigma \alpha_m h_m(x_i))$

adaboost

$$W_{m+1}^{i} = W_{m}^{i} \cdot \beta^{1-|h_m(x_i)-y_i|/2} \qquad e_m = \sum_{y_i=k_m(x_i)} W_m^i \qquad \alpha_m = \frac{e_m}{1-e_m}$$

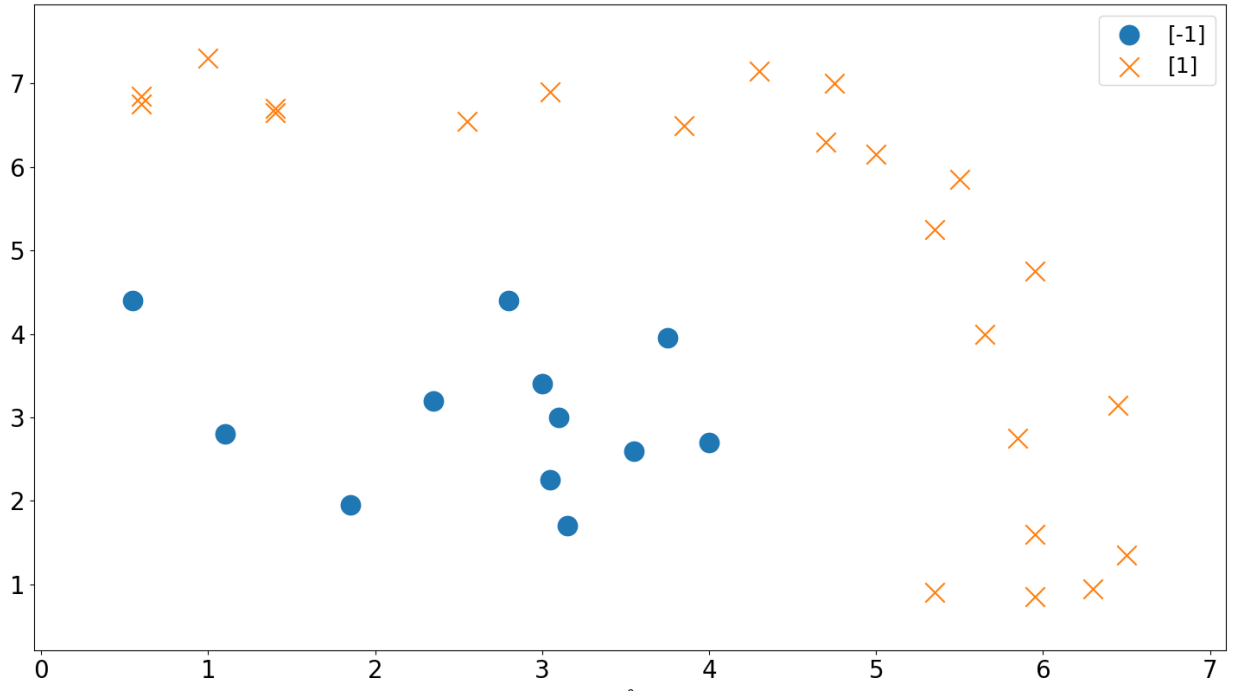
■ 流程图（$y_i = \{-1,1\}$）



■ 简单示例（采用adaboost M1算法）

分析:

- 训练样本是二维坐标，即每个样本有两个特征x和y。
- 使用单层决策树作为弱分类器。
- 为了使结果更直观，两类的标签用-1和1表示。
- 强分类器最后的输出就为： $h_f = \sum_i^n \alpha_i h_i$
- 最后分类： $h_f >= 0, w = 1;$
  $h_f < \ 0, w = -1;$



■ 简单示例

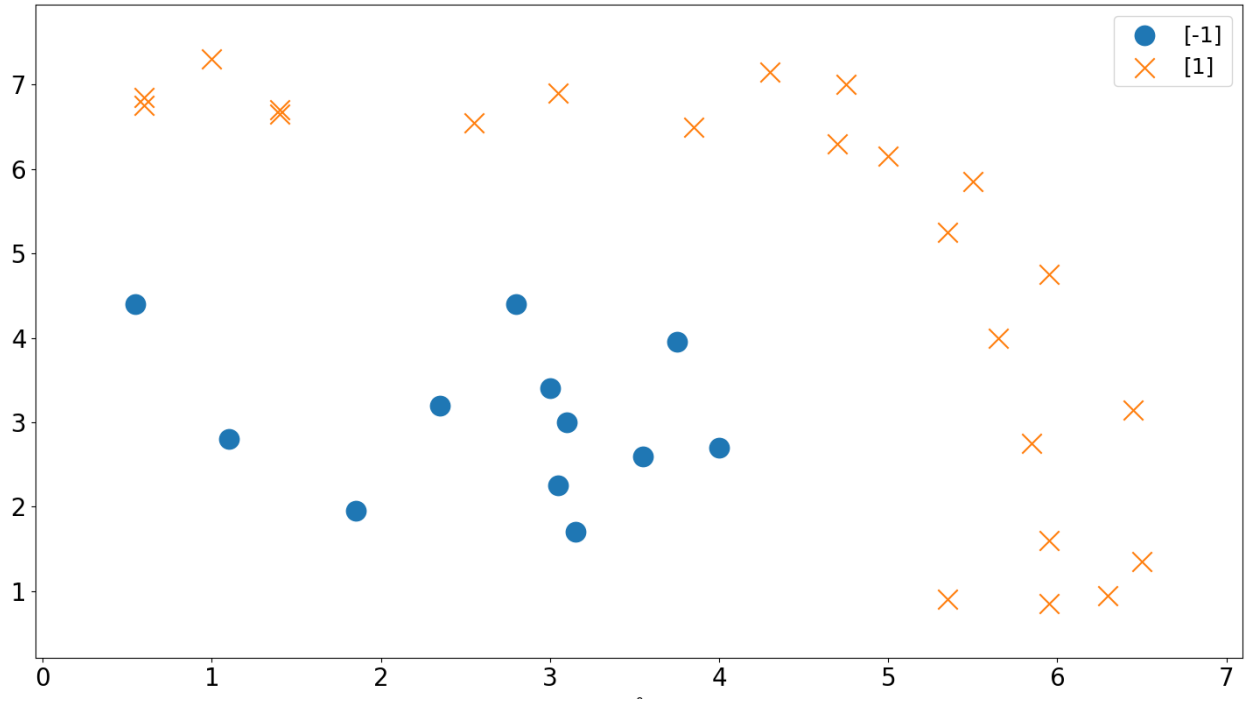
Step 1 – 初始化权重

- 对每个样本初始化一个权重分布：

$$p_1, p_2, \cdots p_N$$

- 归一化权重：

$$w_i = \frac{p_i}{\sum_{j=1}^{N} p_j}$$



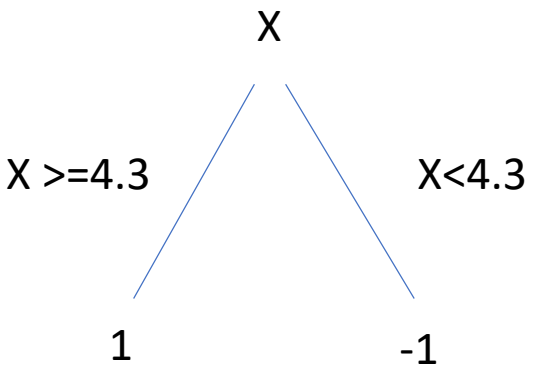
注：如果一开始不知道样本的权重分布情况，则通常开始将所有样本的权重设置成一样。

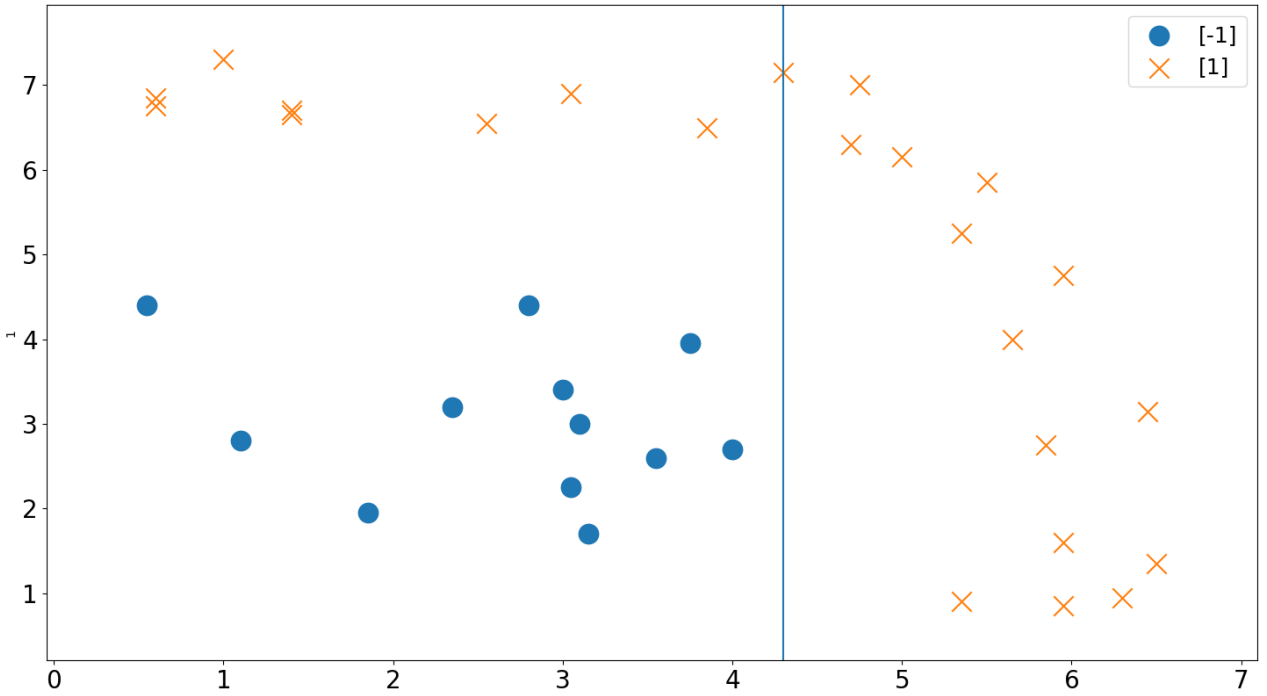
■ 简单示例

Step 2 - 迭代

第一次迭代 – 生成弱分类器:

- 在特征x中找到一个阈值x_t使样本错误率最小。在特征y中找到找到一个阈值y_t使样本错误率最小。最后将错误率最小的特征的阈值作为最后的决策树的阈值。





x_{t1} = 4.3

e_1 = 0.228571

α_1= 0.608198

更新权重到w_2

■ 简单示例

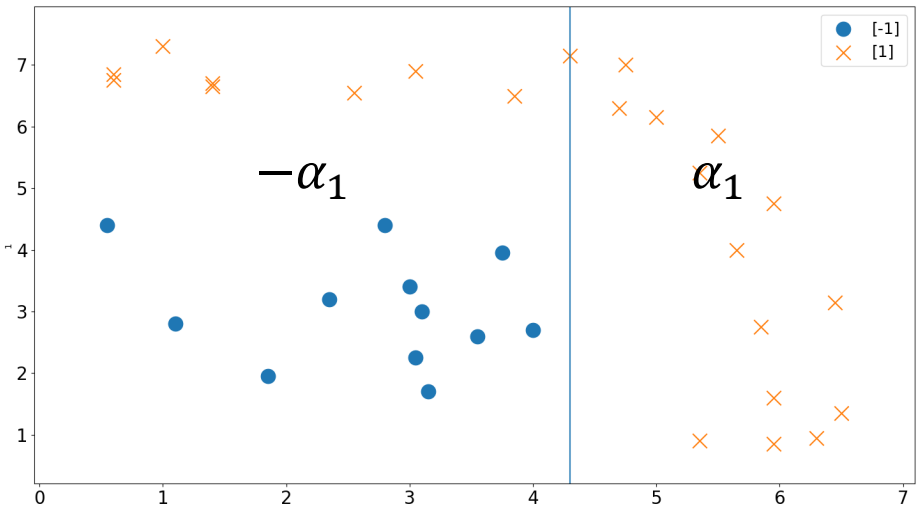
Step 2 - 迭代

第一次迭代 – 生成强分类器:

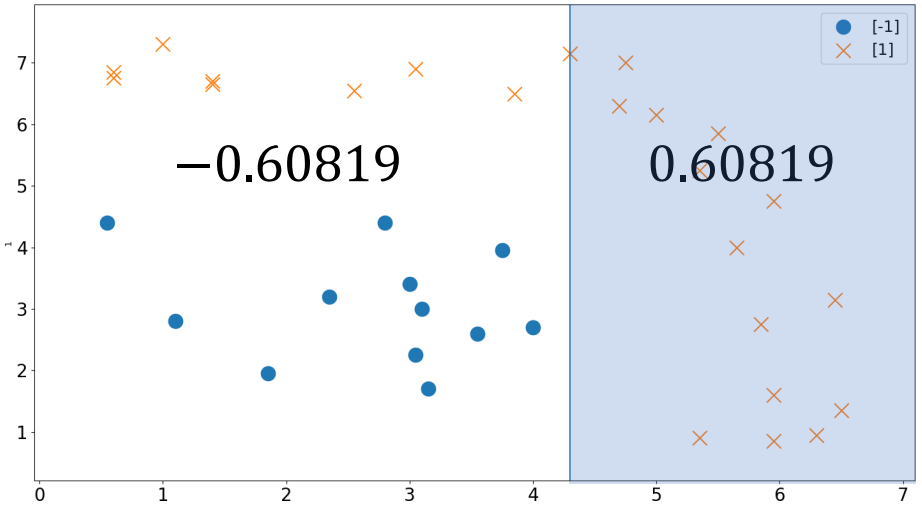
- 第一步只生成了一个弱分类器，则本轮生成的强分类器由一个弱分类器构成，最后的输出:

$$h_f = \sum_i^n \alpha_i h_i$$

- 最后分类: $h_f >= 0, w = 1;$
  $h_f < \ \ 0, w = -1;$



x_{t1} = 4.3
e_1 = 0.228571
α_1 = 0.608198



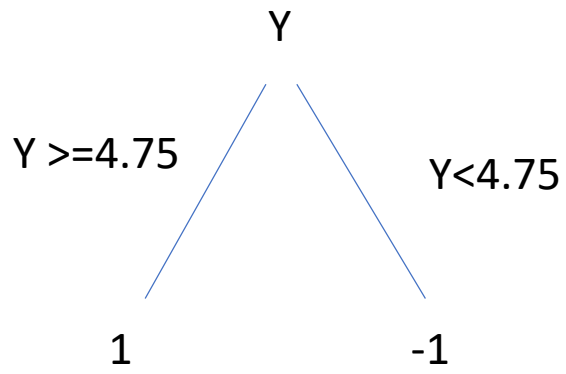
e_f = 0.228571

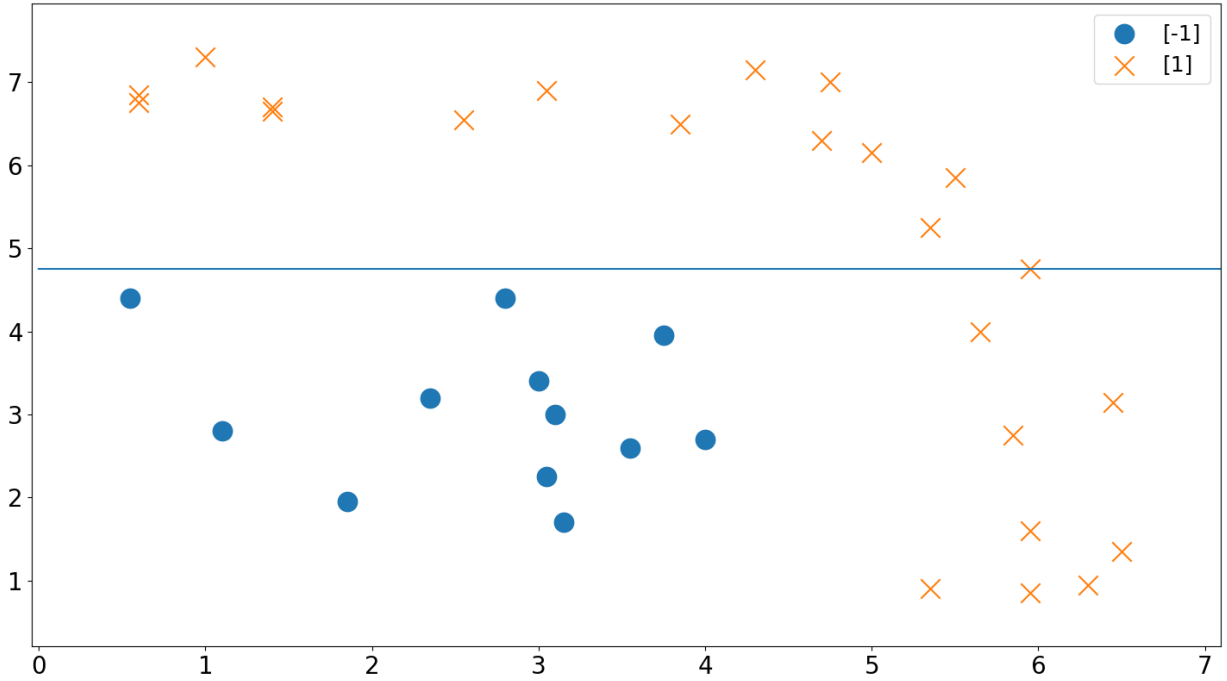
■ 简单示例

Step 2 - 迭代

第二次迭代 – 生成弱分类器:

- 在特征x中找到一个阈值x_t使样本错误率最小。在特征y中找到找到一个阈值y_t使样本错误率最小。最后将错误率最小的特征的阈值作为最后的决策树的阈值。





y_{t1} = 4.75

e_2 = 0.148148

α_2= 0.874600

更新权重到w_3

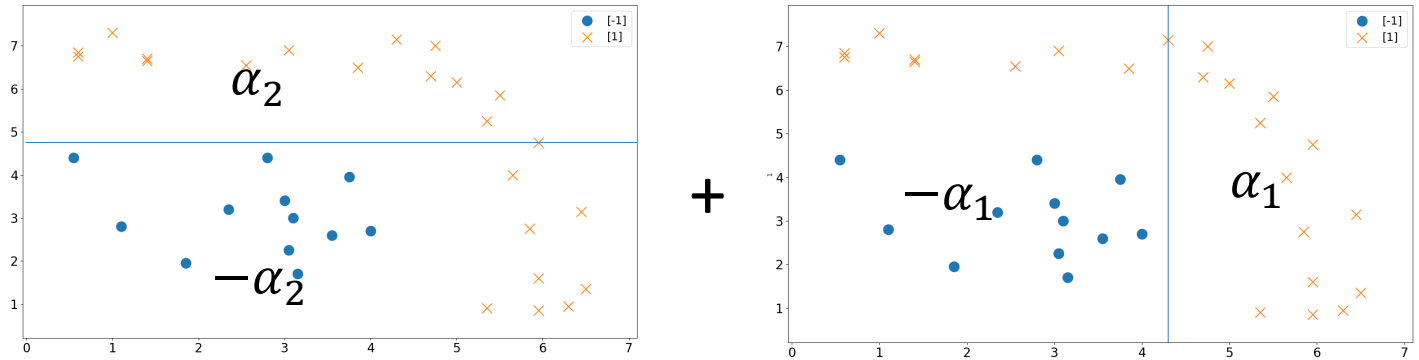
■ 简单示例

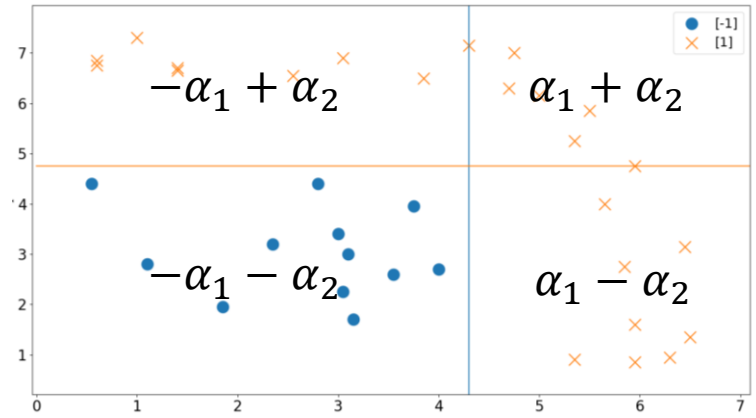
Step 2 - 迭代

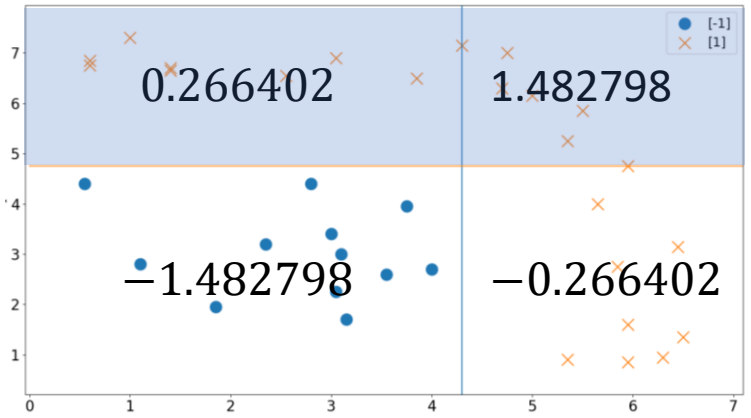
第二次迭代－生成强分类器:

e_2 = 0.148148
α_2= 0.874600

e_1 = 0.228571
α_1= 0.608198







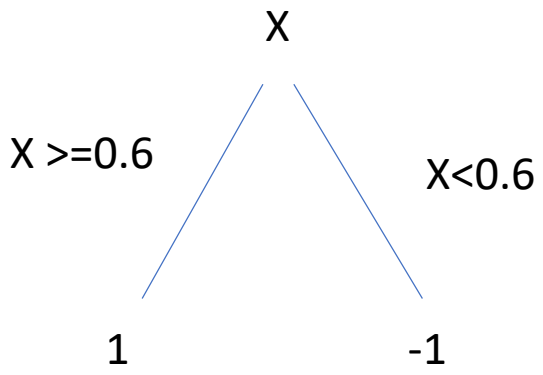
e_f = 0.228571

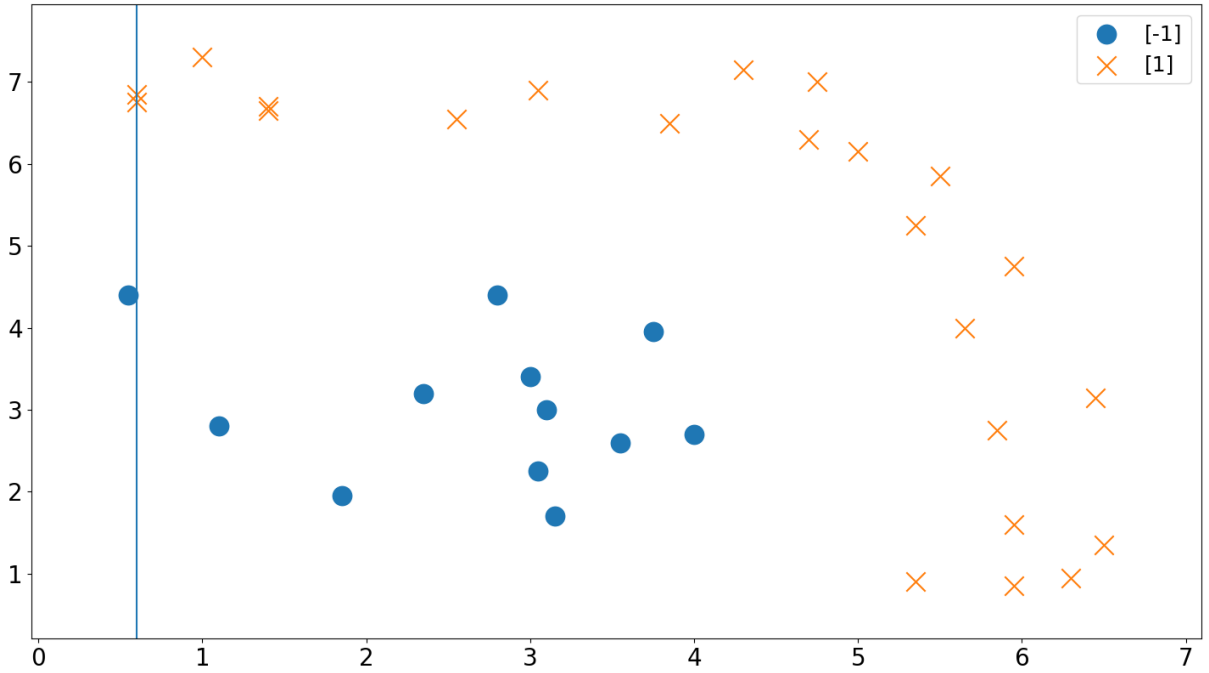
■ 简单示例

Step 2 - 迭代

第三次迭代 – 生成弱分类器:

- 在特征x中找到一个阈值x_t使样本错误率最小。在特征y中找到找到一个阈值y_t使样本错误率最小。最后将错误率最小的特征的阈值作为最后的决策树的阈值。





x_{t2} = 0.6

e_3 = 0.119565

α_3= 0.998277

更新权重到w_4

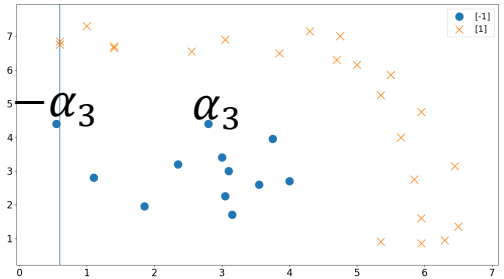
■ 简单示例

e_3 = 0.119565
α_3= 0.998277

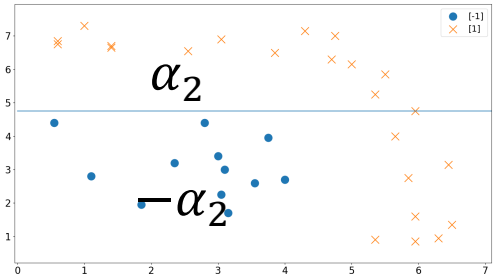
e_2 = 0.148148
α_2= 0.874600

e_1 = 0.228571
α_1= 0.608198

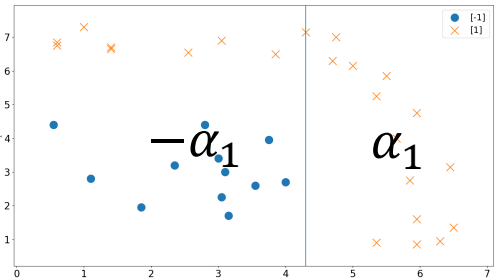
Step 2 - 迭代



+

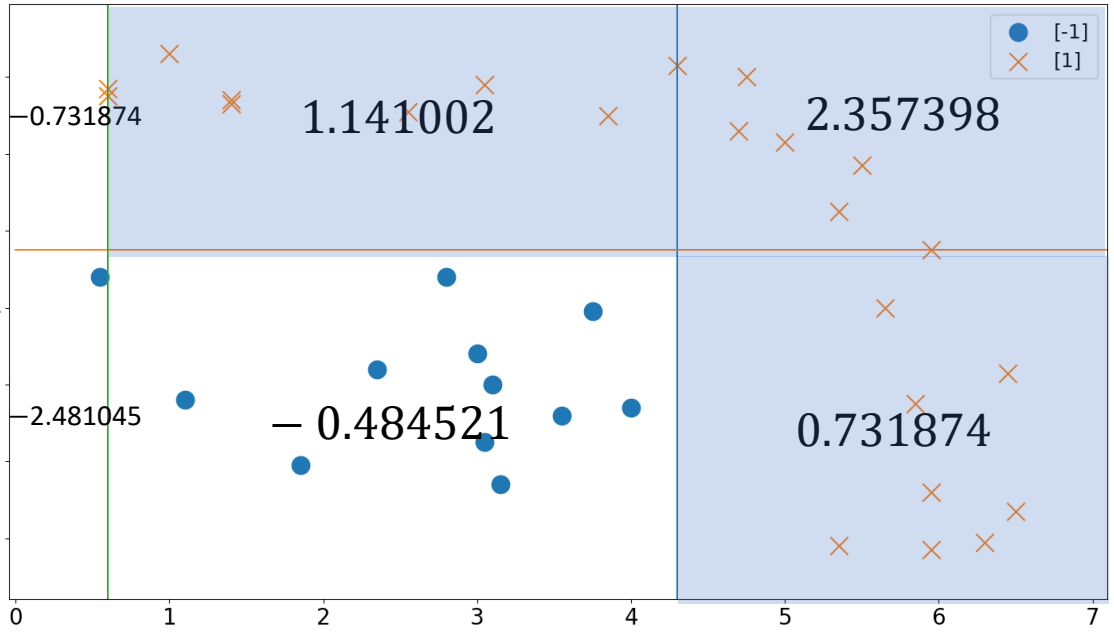


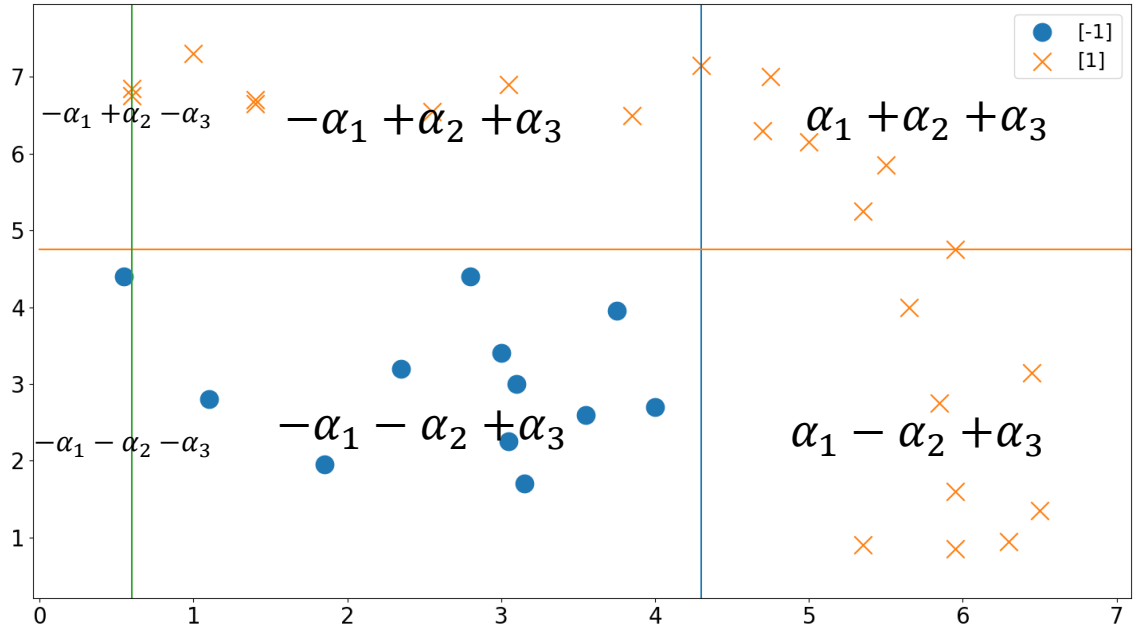
+



第三次迭代 – 生成强分类器:

e_f = 0





adaboost

■ 问题？？？

1、对于二分类单层决策树问题，将导致每次弱分类器的的错误率必小于等于0.5。所以算法才会收敛。如果是多分类问题，若错误率大于0.5则算法结束，为什么？论文算法中有提到。

2、关于二分类的收敛性和性能评估？