第八章：集成学习

By Xian2207, 13689903575, wszhangxian@126.com

1 个体与集成

集成学习（Ensemble Learning）通过构建多个个体学习器并结合起来完成学习任务。分两大类，（1）个体学习器存在强依赖，必须串行生成序列化方法，代表是 Adaptive Boosting（AdaBoost）算法；（2）个体学习器不存在强依赖，可并行生成方法，代表是 Bagging（装袋算法）和 Random Forest（随机森林）。

2 Boosting

Boosting 是一种将弱学习器（weak learner）提升为强学习器的算法。工作机制是：先从初始训练集训练出一个基学习器（base learner），再根据基学习器的表现对训练样本分布进行调整，将基学习器分类错误的样本在后续得到更多关注，基于调整后的样本分布（对训练样本重新采样）来训练下一个学习器，如此重复，直至基学习器数目达到事先指定的值 T，最终将这 T 个基学习器进行加权结合。重复采用的好处是避免基学习器过早学习停止，迭代轮数 T 没达到，导致最终性能不佳。从偏差-方差角度看，Boosting 主要关注减低偏差，因此它可以基于泛华性能较弱的学习器构建成很强的集成。集成规模即集成中包含的个体学习器数目，规模越大，分类边界越复杂，表示能力越强。

Boost 算法中最著名的代表是 AdaBoost（1995），最易理解的就是“加性模型”（additive model）即基学习器的线性组合

$H(x) = \sum_t^T \alpha_t h_t(x)$ 8.4

来最小化指数损失函数（exponential loss function）

$\ell_{exp}(H|D) = E_{x\sim D(e^{-f(x)}H(x))}$ 8.5

算法流程

（1） 初始化训练数据的权值分布

$$y_i \in \{-1,+1\}$$
$$D_1 = w_{11}, w_{12}, \dots, w_{1i}, w_{1N}$$
$$w_{1i} = \frac{1}{N}, i \in [1,2,\dots.,N]$$

（2） 对$m \in [1,2,\dots.,M]$个分类器有

A：使用具有权值分布D_m的训练集学习，得到基本分类器$G_m(x)$

B：计算$G_m(x)$在训练集的分类误差率

$$e_m = P(G_m(x_i \neq y_i)) = \sum_{i=1}^{N} w_{m,i} \cdot I[G_m(x_i) \neq y_i]$$

可看出$G_m(x)$在D_m上的分类错误率是错误分类样本的权值之和。

C：计算$G_m(x)$系数

$$\alpha_m = \frac{1}{2} log \frac{1 - e_m}{e_m}$$

由上式知，分类误差率越小，分类器系数越大，改分类器在最终结果中起的作用越大。故本轮得到的分类器线性组合为：

$$f_m(x) = \sum_{1}^{m} \alpha_m G_m(x)$$

D：更新权值

$$D_{m+1} = w_{m+1,1}, w_{m+1,2}, \ldots, w_{m+1,i}, w_{m+1,N}$$

$$w_{m+1,i} = \begin{cases} \frac{w_{m,i}}{Z_m} e^{-\alpha_m} = \frac{D_m(i)}{2 * e_m}, G(m_i) = y_i \\ \frac{w_{m,i}}{Z_m} e^{\alpha_m} = \frac{D_m(i)}{2 * (1 - e_m)}, G(m_i) \neq y_i \end{cases}$$

$$Z_m = \sum_{i=1}^{N} w_{m,i} \cdot \exp(-\alpha_m y_i G_m(x_i)) = 2\sqrt{e_m(1 - e_m)}$$

Z_m的作用是归一化因子，保证D_{m+1}是一个概率分布，其包含的 N 个权值之和为 1. 可知，当分类正确，数据权值会变小；分类错误，数据权值变大。计算分类误差e_m时，实际计算的是错误分类样本的权值之和。当误差率要求越小，分类器自然会尽量照顾权值大的数据。所以提高本轮那些分类错误样本的权值，使得误分类样本在下一轮起更大作用。之后，分类器将优先考虑它们，努力将其分类正确。

（3） 将 M 个分类器进行线性组合

$$f_m(x) = \sum_{1}^{M} \alpha_m G_m(x)$$

最终分类器为：

$$G(x) = sign(\sum_{1}^{M} \alpha_m G_m(x))$$

注意：boosting 不改变训练集，不断改变每一轮权值，使得训练集数据在基本分类器的学习中起到不同作用。

3 Bagging 与随机森林

Bagging(Bootstrap Aggregating, 1996)是并行集成学习最著名的代表，与标准 AdaBoost 只适用于二分类不同，Bagging 能不经修改地用于多分类、回归等任务。其算法机制是：给定 m 个数量集，我们先随机选出一个样本放入采样集中，再把该样本放回初始数据集，使下次采样该样本仍有可能被选中，这样经过 m 次随机采样操作后，得到一个含有 m 个样本的采样集。如此，可采样出 T 含 m 个训练样本的采样集。基于每个采样集训练出一个基学习器，再将基学习器组合。采样过程中，初始数据集有 63.2%的数据组成采样集，剩下 36.8%的样本可做验证集进行泛化性能的包外估计。当基学习器是决策树时，包外估计 out of bag estimate 有助于决策树进行剪枝处理，当基学习器是神经网络时，包外估计可减小过拟合风险。从偏差-方差角度看，Bagging 关注降低方差。因此它在不剪枝决策树、神经网络等容易受样本扰

动的学习器上效用更加明显。

随机森林（Random Forest, RF）是 Bagging 的一个拓展变体。RF 是以决策树为基学习器构建 Bagging 集成的基础上，进一步在决策树训练过程中引入随机属性选择。传统决策树在选择划分属性时，在当前节点的d个属性集合中选择一个最优属性；而在 RF 中，对基决策树的每个结点，先从该结点的属性集合中随机选择一个包含k个属性的子集。然后再从该子集中选择最优属性用于划分。k控制了随机性的引入程度。若$k = d$，则 RF 退化为传统决策树。通常使用$k = log_2(d)$为推荐值。

RF 简单，易实现，计算开销小。与 Bagging 相比，RF 中基学习器的多样性不仅来自样本扰动还来自属性扰动。而且训练效率常优于 Bagging，因其只需考察一个属性子集。其收敛性与 Bagging 相似，起始性能往往相对较差，因为属性扰动的缘故，个体学习器性能往往有所降低，随着学习器增多，RF 通常会收敛到更低的泛化误差。

4 组合策略

（1） 平均法，数值型输出

方法有两种，简单平均法，即直接计算均值；加权平均法，即各项系数之和为 1，各项≥ 0.研究表明：必须使用非负权重才能确保集成学习性能优于单一个体学习器，因此集中学习一般对个体权重做非负约束。此外，在个体学习器性能相差较大时宜采用加权平均法，性能相近时宜采用简单平均法。

（2） 投票法，分类任务

方法有多种：绝对多数投票法，即某标记过得票过半数的可预测为该类；相对多数投票法，即得票最多的标记可预测为该类；加权投票法，即某些学习器权重大，某些少，最后统计票数最大的；还有类概率，类标记，硬投票等方法。

（3） 学习法

当训练数据很多时，一种更为强大的结合策略是使用学习法，通过另一个学习器来进行结合。

Stacking 是学习法的典型代表。此处将个体学习器称为初级学习器，用于结合的学习器称为次级学习器或者元学习器。Stacking 本身也是一种集成学习方法，也可看作是一种特殊的结合策略。

Stacking 先从初始训练集训练出初级学习器，然后“生成”一个新数据集用于训练次级学习器。新数据集中，初级器的输出被当做样例输入特征，初始样本的标记被当做样例标记。训练阶段，次级训练集是利用初级学习器产生的，若直接用初级学习器的训练集来产生次级训练集，则过拟合风险比较大；因此，一般通过交叉验证或留一法方式，用训练初级学习器未使用的样本生成次级学习器的训练样本。对待测试样本，分别过初、次学习器，即可得到结果。

次级学习器的输入属性表示、学习算法，对 Stacking 集成的泛化性能有很大影响。有研究表明，将初级学习器的输出类概率作为次级学习器的输入属性，用多响应线性回归（Multi-response Linear Regression，MLR）做次级学习算法效果较好。在 MLR 中使用不同属性集效果更佳。

5 多样性

构建泛化能力强的集成学习器，需要的个体学习器应好而不同。理论推导为：集成的泛化误

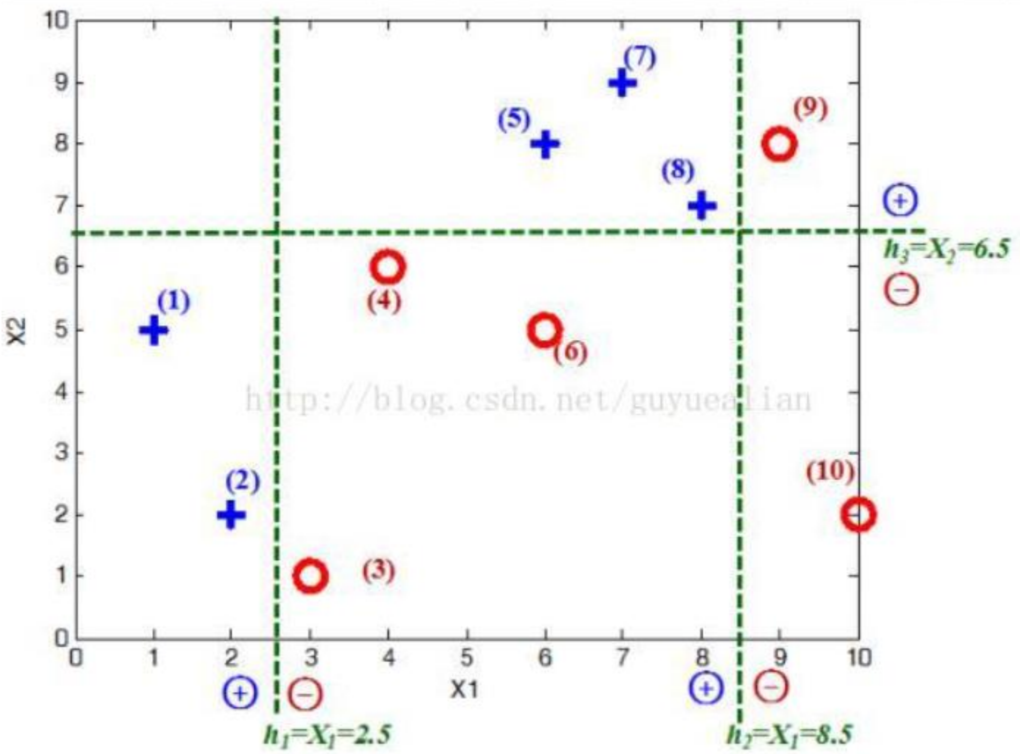
差等于个体学习器泛化误差的加权均值减去个体学习器的加权分歧值。这说明，个体学习器准确性越高，个体学习器泛化误差越小；多样性越大，个体学习器的加权分歧值越大，效果越好。

多样性增强方法有：（1）数据样本扰动，机制是从初始数据集生成不同的子集，利用不同子集训练个体学习器。各个子集可以采用 Bagging 自助采样或 AdaBoost 序列采样。该方法特别适用于决策树，神经网络等训练样本稍加变化就会导致判别模型结果显著变动的学习器。但对稳定学习器，如线性学习器，SVM，朴素贝叶斯，KNN 等不适用；（2）输入属性扰动，机制是不同属性子集代表不同子空间，算法依赖属性的输入。如决策树，随机森林。样本空间大时，随机森林快，样本空间小时，此方法不适宜；（3）输出表示扰动，机制是对输出表示进行操作，改变一些样本的标记，如翻转法；（4）算法参数扰动，机制是对学习器参数设置进行随机设置，如随机森林同时用此法和输入属性扰动。

6 AdaBoost 实例

实例参考链接：https://blog.csdn.net/guyuealian/article/details/70995333 假设样本空间，类的标记如下：

样本序号	1	2	3	4	5	6	7	8	9	10
样本点 X	(1,5)	(2,2)	(3,1)	(4,6)	(6,8)	(6,5)	(7,9)	(8,7)	(9,8)	(10,2)
类别 Y	1	1	-1	-1	1	-1	1	1	-1	-1



数据分析：将这 10 个样本作为训练数据，根据 X 和 Y 的对应关系，可把这 10 个数据分为两类，图中用“+”表示类别 1，用“O”表示类别-1。本例使用水平或者垂直的直线作为分类器，图中已经给出了三个弱分类器，即

$$h_1=\begin{cases}+1, X_1<2.5\\-1, X_1>2.5\end{cases},\quad h_2=\begin{cases}+1, X_1<8.5\\-1, X_1>8.5\end{cases},\quad h_3=\begin{cases}+1, X_2>6.5\\-1, X_2<6.5\end{cases}$$

初始化训练数据的权值分布

$$y_i \in \{-1,+1\}$$
$$D_1 = w_{11}, w_{12}, \ldots, w_{1i}, w_{1N}$$
$$w_{1i} = \frac{1}{N}, i \in [1,2, \ldots., N]$$

故$w_{1i} = \frac{1}{10} = 0.1$，$D_1 = 0.1,\ 0.1,\ 0.1,\ 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1$

第 1 次迭代 t = 1：

A：基本分类器

$$G_m(x)$$

依次考察h_1, h_2, h_3，找到误差最小的。记住：误差率等于该分类器分类错误样本的权值之和。以h_1为标准，小于$X_1(2.5)$即为+1，则错分的有 3 个样本，分别是 5,7,8，故误差率为 0.1+0.1+0.1=0.3；以h_2为标准，小于$X_1(8.5)$即为+1，则错分的有 3 个样本，分别是 3,4,6，故误差率为 0.1+0.1+0.1=0.3；同理，以h_3为标准，,大于$X_1(6.5)$即为正，则错分的有 3 个样本，分别是 1,2,9，故误差率为 0.1+0.1+0.1=0.3.因同为 0.3，那就取第 1 个，即

$$G_m(x) = \begin{cases} +1, X_1 < 2.5 \\ -1, X_1 > 2.5 \end{cases}$$

B：计算基本分类器的误差率

$$e_m = P(G_m(x_i \neq y_i)) = \sum_{i=1}^{N} w_{m,i} \cdot I[G_m(x_i) \neq y_i]$$

第一次分类，$m = 1$，在分类器为h_1的情况下，样本$i = 5,7,8$被错分，故误差率为

$$e_1 = P(G_1(x_i) \neq y_i) = 0.1 + 0.1 + 0.1 = 0.3$$

C：计算$G_m(x)$的系数，即权重

$$\alpha_m = \frac{1}{2} log \frac{1 - e_m}{e_m}$$

故

$$\alpha_1 = \frac{1}{2} * \ln\left(\frac{1 - e_1}{e_1}\right) = 0.5 * \ln\left(\frac{1 - 0.3}{0.3}\right) = 0.4236$$

α_1代表h_1在最终分类函数中所占的权重为 0.4236，故本轮得到的分类器组合为

$$f_m(x) = \sum_{1}^{m} \alpha_m G_m(x) = f_1(x) = 0.4236 * h_1(x)$$

D：更新权重

$$D_{m+1} = w_{m+1,1}, w_{m+1,2}, \ldots, w_{m+1,i}, w_{m+1,N}$$
$$w_{m+1,i} = \begin{cases} \frac{D_m(i)}{2 * (1 - e_m)}, G(m_i) = y_i \\ \frac{D_m(i)}{2 * e_m}, G(m_i) \neq y_i \end{cases}$$

第 1 次正确分类样本：1,2,3,4,6,9,10,共 7 个，故

$$w_{1+1,1} = \frac{D_m(i)}{2 * e_m} = \frac{D_1(1)}{2 * (1 - e_1)} = \frac{0.1}{2 * 0.7} = \frac{1}{14}$$
$$w_{1+1,2} = \frac{D_m(i)}{2 * e_m} = \frac{D_1(2)}{2 * (1 - e_1)} = \frac{0.1}{2 * 0.7} = \frac{1}{14}$$

$$w_{1+1,3}=\frac{D_m(i)}{2*e_m}=\frac{D_1(3)}{2*(1-e_1)}=\frac{0.1}{2*0.7}=\frac{1}{14}$$
$$w_{1+1,4}=\frac{D_m(i)}{2*e_m}=\frac{D_1(4)}{2*(1-e_1)}=\frac{0.1}{2*0.7}=\frac{1}{14}$$
$$w_{1+1,6}=\frac{D_m(i)}{2*e_m}=\frac{D_1(6)}{2*(1-e_1)}=\frac{0.1}{2*0.7}=\frac{1}{14}$$
$$w_{1+1,9}=\frac{D_m(i)}{2*e_m}=\frac{D_1(9)}{2*(1-e_1)}=\frac{0.1}{2*0.7}=\frac{1}{14}$$
$$w_{1+1,10}=\frac{D_m(i)}{2*e_m}=\frac{D_1(10)}{2*(1-e_1)}=\frac{0.1}{2*0.7}=\frac{1}{14}$$

第 1 次错误分类样本：5,7,8,共 3 个，故

$$w_{1+1,5}=\frac{D_m(i)}{2*e_m}=\frac{D_1(5)}{2*e_1}=\frac{0.1}{2*0.3}=\frac{1}{6}$$
$$w_{1+1,7}=\frac{D_m(i)}{2*e_m}=\frac{D_1(7)}{2*e_1}=\frac{0.1}{2*0.3}=\frac{1}{6}$$
$$w_{1+1,8}=\frac{D_m(i)}{2*e_m}=\frac{D_1(8)}{2*e_1}=\frac{0.1}{2*0.3}=\frac{1}{6}$$

更新后的权值分布

$$D_1=\frac{1}{14},\frac{1}{14},\frac{1}{14},\frac{1}{14},\frac{1}{6},\frac{1}{14},\frac{1}{6},\frac{1}{6},\frac{1}{14},\frac{1}{14}$$

注意：更新后正确分类的样本权值由 0.1 减小到$\frac{1}{14}$，错误分类的样本由 0.1 增大为$\frac{1}{6}$。

第 2 次迭代 t = 2：

A：基本分类器

$$G_2(x)$$

再依次考察h_1,h_2,h_3，找到误差最小的。记住：误差率等于该分类器分类错误样本的权值之和。以h_1为标准，小于$X_1(2.5)$即为+1，则错分的有 3 个样本，分别是 5,7,8，故误差率为$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}$；以$h_2$为标准，小于$X_1(8.5)$即为+1，则错分的有 3 个样本，分别是 3,4,6，故误差率为$\frac{1}{14}+\frac{1}{14}+\frac{1}{14}=\frac{3}{14}$；同理，以$h_3$为标准，,大于$X_1(6.5)$即为正，则错分的有 3 个样本，分别是 1,2,9，故误差率为为$\frac{1}{14}+\frac{1}{14}+\frac{1}{14}=\frac{3}{14}$.故$h_2=h_3$，那就取$h_2$，即

$$G_2(x)=\{\begin{matrix}+1, X_1<8.5\\-1, X_1>8.5\end{matrix}$$

B：计算基本分类器的误差率

$$e_m=P(G_m(x_i\neq y_i))=\sum_{i=1}^{N}w_{m,i}\cdot I[G_m(x_i)\neq y_i]$$

在分类器为h_2的情况下，样本$i=3,4,6$被错分，故误差率为

$$e_2=P(G_2(x_i)\neq y_i)=\frac{1}{14}+\frac{1}{14}+\frac{1}{14}=\frac{3}{14}$$

C：计算$G_m(x)$的系数，即权重

$$\alpha_m = \frac{1}{2} log \frac{1-e_m}{e_m}$$

故

$$\alpha_2 = \frac{1}{2} * \ln\left(\frac{1-e_2}{e_2}\right) = 0.5 * \ln\left(\frac{1-\frac{3}{14}}{\frac{3}{14}}\right) = 0.6496$$

α_2代表h_2在最终分类函数中所占的权重为 0.6496，故本轮得到的分类器组合为

$$f_m(x) = \sum_{1}^{m} \alpha_m G_m(x) = f_2(x) = 0.4236 * h_1(x) + 0.6496 * h_2(x)$$

D：更新权重

$$D_{m+1} = w_{m+1,1}, w_{m+1,2}, \ldots, w_{m+1,i}, w_{m+1,N}$$

$$w_{m+1,i} = \begin{cases} \dfrac{D_m(i)}{2*(1-e_m)}, G(m_i) = y_i \\ \dfrac{D_m(i)}{2*e_m}, G(m_i) \neq y_i \end{cases}$$

第 2 次正确分类样本：1,2,5,7,8,9,10,共 7 个，故

$$w_{2+1,1} = \frac{D_m(i)}{2*e_m} = \frac{D_2(1)}{2*(1-e_2)} = \frac{\frac{1}{14}}{2*\frac{11}{14}} = \frac{1}{22}$$

$$w_{2+1,2} = \frac{D_m(i)}{2*e_m} = \frac{D_2(2)}{2*(1-e_2)} = \frac{\frac{1}{14}}{2*\frac{11}{14}} = \frac{1}{22}$$

$$w_{2+1,5} = \frac{D_m(i)}{2*e_m} = \frac{D_2(5)}{2*(1-e_2)} = \frac{\frac{1}{6}}{2*\frac{11}{14}} = \frac{7}{66}$$

$$w_{2+1,7} = \frac{D_m(i)}{2*e_m} = \frac{D_2(7)}{2*(1-e_2)} = \frac{\frac{1}{6}}{2*\frac{11}{14}} = \frac{7}{66}$$

$$w_{2+1,8} = \frac{D_m(i)}{2*e_m} = \frac{D_2(8)}{2*(1-e_2)} = \frac{\frac{1}{6}}{2*\frac{11}{14}} = \frac{7}{66}$$

$$w_{2+1,9} = \frac{D_m(i)}{2*e_m} = \frac{D_2(9)}{2*(1-e_2)} = \frac{\frac{1}{14}}{2*\frac{11}{14}} = \frac{1}{22}$$

$$w_{2+1,10} = \frac{D_m(i)}{2*e_m} = \frac{D_2(10)}{2*(1-e_2)} = \frac{\frac{1}{14}}{2*\frac{11}{14}} = \frac{1}{22}$$

第 2 次错误分类样本：3,4,6,共 3 个，故

$$w_{2+1,3} = \frac{D_m(i)}{2*e_m} = \frac{D_1(3)}{2*e_2} = \frac{\frac{1}{14}}{2*\frac{3}{14}} = \frac{1}{6}$$

$$w_{2+1,4}=\frac{D_m(i)}{2*e_m}=\frac{D_1(4)}{2*e_2}=\frac{\frac{1}{14}}{2*\frac{3}{14}}=\frac{1}{6}$$

$$w_{2+1,6}=\frac{D_m(i)}{2*e_m}=\frac{D_1(6)}{2*e_2}=\frac{\frac{1}{14}}{2*\frac{3}{14}}=\frac{1}{6}$$

更新后的权值分布变为

$$D_3=\frac{1}{22},\frac{1}{22},\frac{1}{6},\frac{1}{6},\frac{7}{66},\frac{1}{6},\frac{7}{66},\frac{7}{66},\frac{1}{22},\frac{1}{22}$$

第 3 次迭代 t = 3：

A：基本分类器

$$G_3(x)$$

再依次考察h_1,h_2,h_3，找到误差最小的。记住：误差率等于该分类器分类错误样本的权值之和。以h_1为标准，小于$X_1(2.5)$即为+1，则错分的有 3 个样本，分别是 5,7,8，故误差率为$\frac{7}{66}+\frac{7}{66}+\frac{7}{66}=\frac{7}{22}$；以$h_2$为标准，小于$X_1(8.5)$即为+1，则错分的有 3 个样本，分别是 3,4,6，故误差率为$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}$；同理，以$h_3$为标准，,大于$X_1(6.5)$即为正，则错分的有 3 个样本，分别是 1,2,9，故误差率为为$\frac{1}{22}+\frac{1}{22}+\frac{1}{22}=\frac{3}{22}$.故$h_3$最小，取$h_3$，即

$$G_3(x)=\begin{cases}+1, X_1>6.5\\-1, X_1<6.5\end{cases}$$

B：计算基本分类器的误差率

$$e_m=P(G_m(x_i\neq y_i))=\sum_{i=1}^{N}w_{m,i}\cdot I[G_m(x_i)\neq y_i]$$

在分类器为h_3的情况下，样本$i=1,2,9$被错分，故误差率为

$$e_3=P(G_3(x_i)\neq y_i)=\frac{1}{22}+\frac{1}{22}+\frac{1}{22}=\frac{3}{22}$$

C：计算$G_m(x)$的系数，即权重

$$\alpha_m=\frac{1}{2}log\frac{1-e_m}{e_m}$$

故

$$\alpha_2=\frac{1}{2}*\ln\left(\frac{1-e_2}{e_2}\right)=0.5*\ln\left(\frac{1-\frac{3}{22}}{\frac{3}{22}}\right)=0.9229$$

α_2代表h_2在最终分类函数中所占的权重为 0.9229，故本轮得到的分类器组合为

$$f_m(x)=\sum_{1}^{m}\alpha_m G_m(x)=f_3(x)=0.4236*h_1(x)+0.6496*h_2(x)+0.9229*h_3(x)$$

D：更新权重

$$D_{m+1}=w_{m+1,1},w_{m+1,2},\ldots,w_{m+1,i},w_{m+1,N}$$

$$w_{m+1,i} = \begin{cases} \frac{D_m(i)}{2*(1-e_m)}, G(m_i) = y_i \\ \frac{D_m(i)}{2*e_m}, G(m_i) \neq y_i \end{cases}$$

第 3 次正确分类样本：3,4,5,6,7,8,10,共 7 个，故

$$w_{3+1,1} = \frac{D_m(i)}{2*e_m} = \frac{D_3(3)}{2*(1-e_3)} = \frac{\frac{1}{6}}{2*\frac{19}{22}} = \frac{11}{114}$$

$$w_{3+1,3} = \frac{D_m(i)}{2*e_m} = \frac{D_2(4)}{2*(1-e_3)} = \frac{\frac{1}{6}}{2*\frac{19}{22}} = \frac{11}{114}$$

$$w_{3+1,4} = \frac{D_m(i)}{2*e_m} = \frac{D_2(5)}{2*(1-e_3)} = \frac{\frac{7}{66}}{2*\frac{19}{22}} = \frac{7}{114}$$

$$w_{3+1,5} = \frac{D_m(i)}{2*e_m} = \frac{D_2(6)}{2*(1-e_3)} = \frac{\frac{1}{6}}{2*\frac{19}{22}} = \frac{11}{114}$$

$$w_{3+1,7} = \frac{D_m(i)}{2*e_m} = \frac{D_2(7)}{2*(1-e_3)} = \frac{\frac{7}{66}}{2*\frac{19}{22}} = \frac{7}{114}$$

$$w_{3+1,8} = \frac{D_m(i)}{2*e_m} = \frac{D_2(8)}{2*(1-e_3)} = \frac{\frac{7}{66}}{2*\frac{19}{22}} = \frac{7}{114}$$

$$w_{3+1,10} = \frac{D_m(i)}{2*e_m} = \frac{D_2(10)}{2*(1-e_3)} = \frac{\frac{1}{22}}{2*\frac{19}{22}} = \frac{1}{38}$$

第 3 次错误分类样本：1,2,9,共 3 个，故

$$w_{3+1,3} = \frac{D_m(i)}{2*e_m} = \frac{D_1(3)}{2*e_2} = \frac{\frac{1}{22}}{2*\frac{3}{22}} = \frac{1}{6}$$

$$w_{3+1,4} = \frac{D_m(i)}{2*e_m} = \frac{D_1(4)}{2*e_2} = \frac{\frac{1}{22}}{2*\frac{3}{22}} = \frac{1}{6}$$

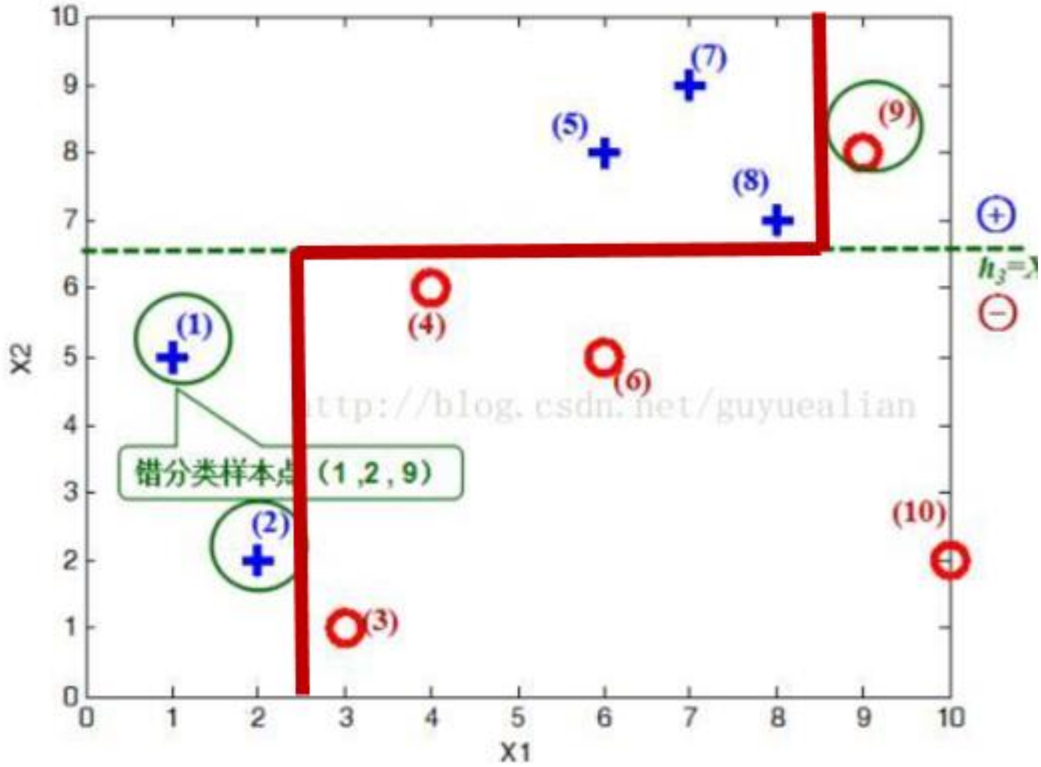
$$w_{3+1,6} = \frac{D_m(i)}{2*e_m} = \frac{D_1(6)}{2*e_2} = \frac{\frac{1}{22}}{2*\frac{3}{22}} = \frac{1}{6}$$

更新后的权值分布变为

$$D_4 = \frac{1}{6}, \frac{1}{6}, \frac{11}{114}, \frac{11}{114}, \frac{7}{114}, \frac{11}{114}, \frac{7}{114}, \frac{7}{114}, \frac{1}{6}, \frac{1}{38}$$

最终最强分类器及其对应的图示（红色线段即为$f_3(x)$）

$$f_3(x) = 0.4236 * h_1(x) + 0.6496 * h_2(x) + 0.9229 * h_3(x)$$



优点：

（1） AdaBoost 提供的是一种框架，框架内的子分类器如h_1, h_2, h_3子分类器方法可任意选择，如以h_1为标准，小于$X_1(2)$或h_1决策树，h_2朴素贝叶斯，h_3为 SVM 等。AdaBoost 不用对特征进行筛选，也不存在过拟合现象；

（2） 不需要类似朴素贝叶斯分类器需要的先验概率，最后得到的强分类器精度依赖于所有弱分类器，不论人造还是真实数据，AdaBoost 都能提高学习精度，即不用知道错误率上线，最终最强分类器可以自己调整；

缺点：

（1）训练时使难以分类的样本权值成指数增长，训练会偏向这些难分类样本，导致训练时间过长，最强分类器难以得到，即分类器组合容易受噪声干扰；

（2）强分类器依赖于弱分类器，而弱分类器训练时间往往过长。