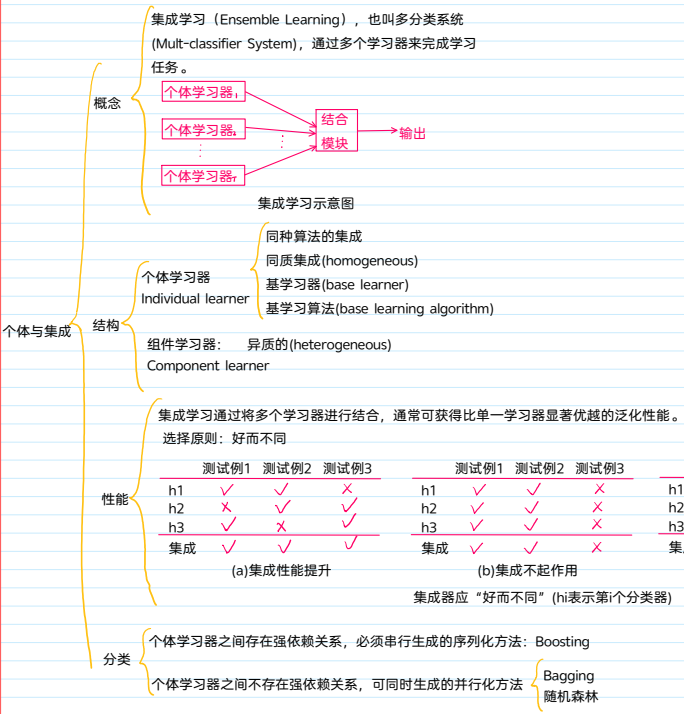


集成学习



Boosting

介绍

AdaBoost

二分类举例

给定如下表所示训练数据。假设个体学习器由x和y产生，其阈值v使该分类器在训练数据集上分类误差率最低。

我们首先认为xi(i=1,2,...,10)的权重是一样的，即每一个数据同等重要（权重用于计算误差）

在权值分布为D1的训练数据集上，阈值v取2.5时分类误差最低（此时x=6,7,8被错分为反例，误差为它们的权重之和e1=0.1+0.1+0.1=0.3，误差率在小于1/2时才有意义），故个体学习器 $G(x) = \begin{cases} 1, & x \leq 2.5 \\ -1, & x > 2.5 \end{cases}$

$\alpha_i = \frac{1}{2} \ln \frac{1 - e_i}{e_i}$ ，可发现只有 $e_i < 0.5$ 时， α_i 是正的，个体学习器才是有效的

根据误差e1计算系数 $\alpha_1 = 0.4236$ ，更新训练数据的权值分布

$w_{m+1,i} = \frac{w_m \exp(-\alpha_m y_i G_m(x_i))}{Z_m}, i=1,2,...,N$

$Z_m = \sum_{i=1}^N w_{m,i} \exp(-\alpha_m y_i G_m(x_i))$ ， Z_m 是为验证每次数值是否和为1

通过指数损失函数exp(x)调整权重，分类正确的降低权重(yi和Gm(xi)同号则yiGm(xi)>0,- $\alpha_m y_i G_m(x_i)$ <0，分类错误的增加权重)：