base_estimator: 默认为DecisionTreeClassifier。

algorithm:默认为SAMME.R,迭代一般比SAMME快,但需要上面的弱分类学习器还需要支持概率预测,也就是在scikit-learn中弱分类学习器对应的预测方法除了predict还需要有predict_proba

n_estimators: 迭代次数,默认为50

learning_rate: 学习率,默认为1.0。

random_state: 随机数生成器

Adaboost

sklearn.ensemble.AdaBoostClassifier

AdaBoost就是损失函数为指数损失的Boosting算法。

总概

Boosting算法要涉及到两个部分,加法模型和前向分步算法。加法模型就是说强分类器由一系列弱分类器线性相加而成,前向分步就是说在训练过程中,下一轮迭代产生的分类器是在上一轮的基础上训练得来的。

初始化训练数据(每个样本)的权值分布:如果有N个样本,则每一个训练的样本点最开始时都被赋予相同的权重: 1/N。

训练弱分类器。具体训练过程中,如果某个样本已经被准确地分类,那么在构造下一个训练集中,它的权重就被降低;相反,如果某个样本点没有被准确地分类,那么它的权重就得到提高。同时,得到弱分类器对应的话语权。然后,更新权值后的样本集被用于训练下一个分类器,整个训练过程如此迭代地进行下去。

原理

将各个训练得到的弱分类器组合成强分类器。各个弱分类器的训练过程结束后,分类误差率小的弱分类器的话语权较大,其在最终的分类函数中起着较大的决定作用,而分类误差率大的弱分类器的话语权较小,其在最终的分类函数中起着较小的决定作用。换言之,误差率低的弱分类器在最终分类器中占的比例较大,反之较小。