

PR homework10

李逸思 自动化系 2016310707

AdaBoosting programming

2.1 Decision stump 算法简述如下:

对于 $j=1:p$,

Step1 对矩阵 X 的第 j 列按照从小到大的顺序进行排序, 并根据 X_j 的排序将 y 和 w 对应排序

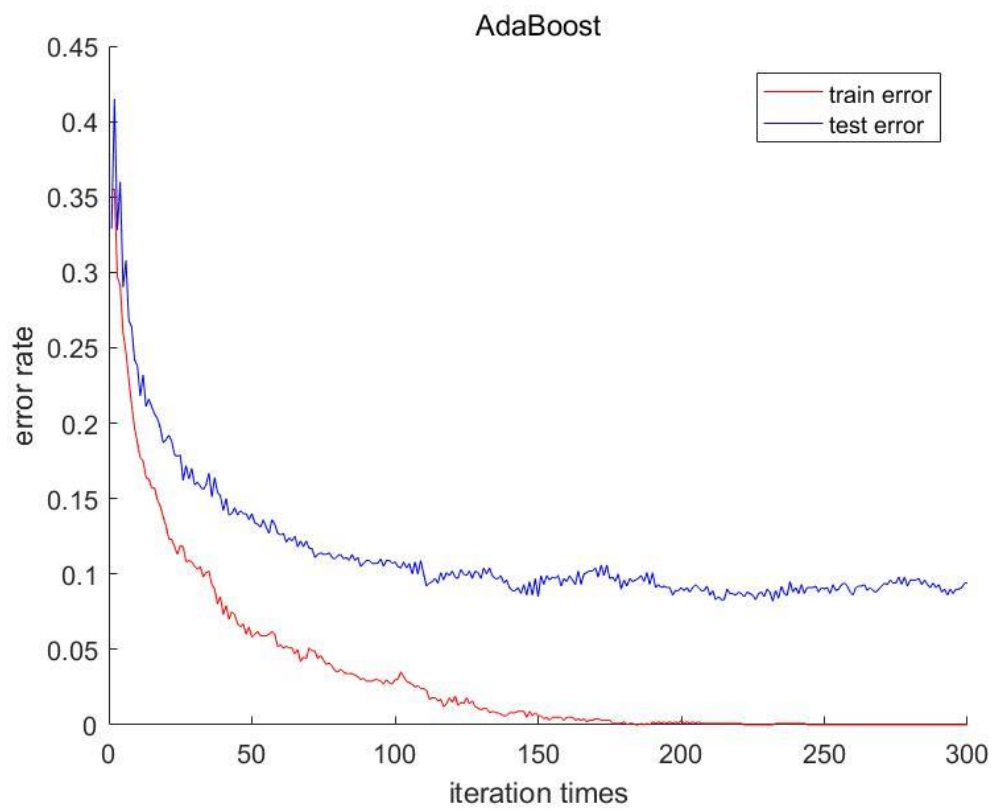
Step2 求 $y.*w$ 的累加和最大值 ma 和最小值 mi

Step3 若 $\text{abs}(ma) > \text{abs}(mi)$, 则 $d_j=1$, 阈值 a_j 对应最大值 ma ; 否则 $d_j=-1$, 阈值 a_j 对应最小值 mi

算法 matlab 代码实现见附件, 其中循环次数为 p , 每次循环使用 `sort` 函数, `sort` 函数时间复杂度为 $O(n \log n)$, 故 Decision stump 算法复杂度为 $O(pn \log n)$ 。

2.2 完成权重更新及错误率计算的 matlab 代码见附件。

2.3 迭代 300 次, 得到 adaboost 算法在测试数据和训练数据上的错误率如图一, 从图中可见, 迭代 150 次后 adaboost 算法在测试集和训练集上的正确率都趋于稳定, 最终, adaboost 在测试数据上的正确率趋于稳定并接近 100%, 在训练数据上的正确率趋于稳定约为 90%。说明 adaboost 是一种分类正确率较高且稳定性较强的性能良好的分类算法。



图一 adaboost 正确率