MONASH University

ETC3250 Business Analytics: Advanced Classification - Boosting

Souhaib Ben Taieb, Di Cook, Rob Hyndman

October 5, 2015

Overview

- Expand the data by giving more weight to the troublesome cases.
- Intuitively, this could be thought of as increasing the amount of data set near the boundaries between the groups.
- Early work description at `http://www.boosting.org`, and see adaboost algorithm

Algorithm

1. Input: $L = (x_i, y_i), i = 1, ..., n, y_i \in \{1, ..., g\}$. Set $r_i = y_i$, and $\hat{y} = 0$.
2. For $b = 1, 2, ..., B$:

   - Fit $\hat{y}^b$ to (x, r).
   - Update $\hat{y} \leftarrow \hat{y} + \lambda \hat{y}^b$.
   - Update the residuals $r_i \leftarrow r_i + \lambda \hat{y}^b$.

3. Output: boosted model, $\hat{y} = \sum_{b=1}^{B} \lambda \hat{y}^b$.

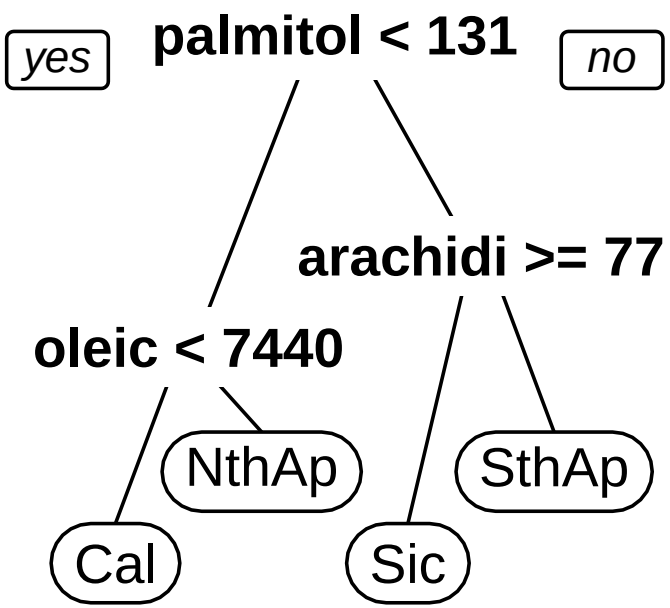
Example

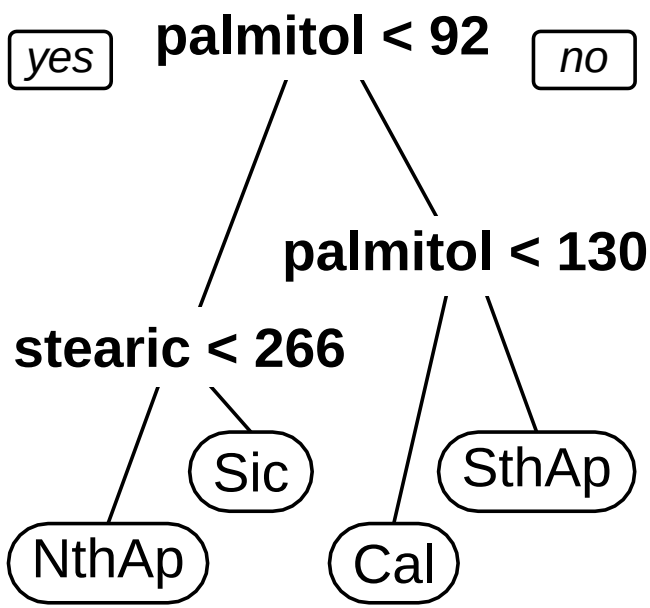
```
olive.adaboost <- boosting(area~., data=olive.s.tr, control=rp
table(olive.s.tr$area, olive.adaboost$class); table(olive.s.ts
```

```
##
##           Cal NthAp Sic SthAp
##    Cal     28     0   0     0
##    NthAp    0    13   0     0
##    Sic      0     0  18     0
##    SthAp    0     0   0   102
##
##           Cal NthAp Sic SthAp
##    Cal     27     0   0     1
##    NthAp    1     7   3     1
##    Sic      5     2   8     3
##    SthAp    2     0   1   101
```

Example

```
## [1] 16 34
```





Examine how it performs with different number of steps