Machine Learning HW8

Lina

December 24, 2016

1 Understanding Decision Tree

(i)The formula to compute the entropy is:

$$E(D) = -\sum_{j=1}^{J} p_j log p_j$$

so,we can easily get the original Information entropy is $E(D) = -\frac{9}{14}log\frac{9}{14} - \frac{5}{14}log\frac{5}{14} = 0.9403$ then we compute the Information Gain according to one attribute with the following fomula:

$$G(D, A) = E(D) - \sum_{i=1}^{n} \frac{N_i}{N} E(D_i)$$

where A represents the attribute.

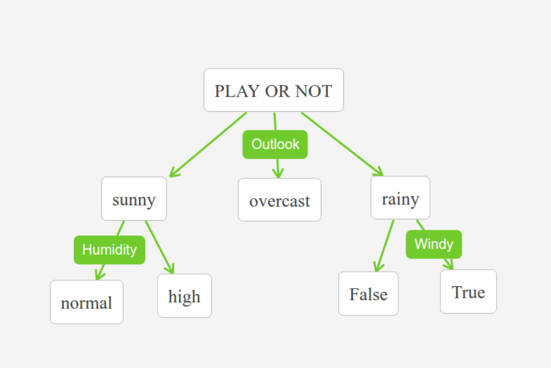
$$G(D, outlook) = E(D) - 0.6935 = 0.2468$$

$$G(D, temperature) = E(D) - 0.9111 = 0.0292$$

$$G(D, humidity) = E(D) - 0.7884 = 0.1519$$

$$G(D, windy) = E(D) - 0.8922 = 0.0481$$

Outlook's Information Gain is the largest,so we choose it as the standard to classify the data.Through the same way we can get the ID3 tree as shown in the image.



(ii)Using Information Gain Ratio, we use the fomula:

$$GR(D,A)=\frac{G(D,A)}{SI(D,A)}$$

where $SI(D,A)=-\sum_{i=1}^{n}\frac{N_i}{N}log\frac{N_i}{N}$

$$SI(D,outlook)=1.5774$$

$$SI(D,temperature)=1.5567$$

$$SI(D,humidity)=1$$

$$SI(D,windy)=0.9852$$

then we can get the $GR(D,A)$ for every attribute:

$$GR(D,outlook)=0.1565$$

$$GR(D,temperature)=0.01876$$

$$GR(D,humidity)=0.1519$$

$$GR(D,windy)=0.0488$$

so we choose outlook to split for the first step.

(iii)Using the Gini Index ,we need to use the fomula:

$$Gini(D)=1-\sum_{i=1}^{m}p_i^2$$

$$Gini(D,R)=\sum_{j=1}^{N}\frac{|D_j|}{|D|}Gini(D_j)$$

through computation,we can get that,

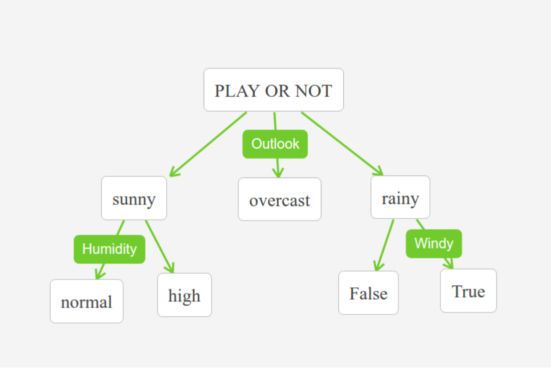
$$Gini(D,outlook)=0.3429$$

$$Gini(D,temperature)=0.4405$$

$$Gini(D,humidity)=0.3673$$

$$Gini(D,windy)=0.4286$$

so we choose outlook to split for the first step , in the same way ,we can get the Cart tree.And finally it is the same with ID3 Tree.



2 Learning Ensemble Methods

Using the Boosting framework to describe the l_2-Boosting algorithm and the loss function is,

$$l(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$$

1. Initialize $f_0(x) = 0$ 2. For $m = 1$ to M:
(1) First we compute the gradient of the loss function with respect to f

$$(g_m)_i = \frac{\partial}{\partial f} l[y_i, f(x_i)]|_{f=f(x_i)}$$

$$= \hat{y} - y|_{f=f(x_i)}$$

and get g_m
(2) Fit regression model to negative gradient $-g_m$:

$$p_m = arg \min_{p \in F} \sum_{i=1}^{N} [-(g_m)_i - p(x_i)]^2$$

(3)Choose fixed stepsize $v_m = v \in (0, 1]$, or take a "line search strategy"

$$v_m = arg \min_{v>0} \sum i = 1^N l[y_i, f_{m-1}(x_i + vp_m(x_i))]$$

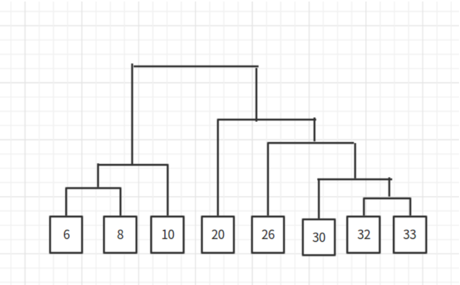
(4) Take this step

$$f_m(x) = f_{m-1}(x) + v_m p_m(x)$$

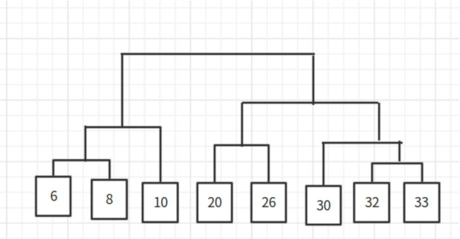
3 Clustering

3.1 Hierarchical clustering

(a)



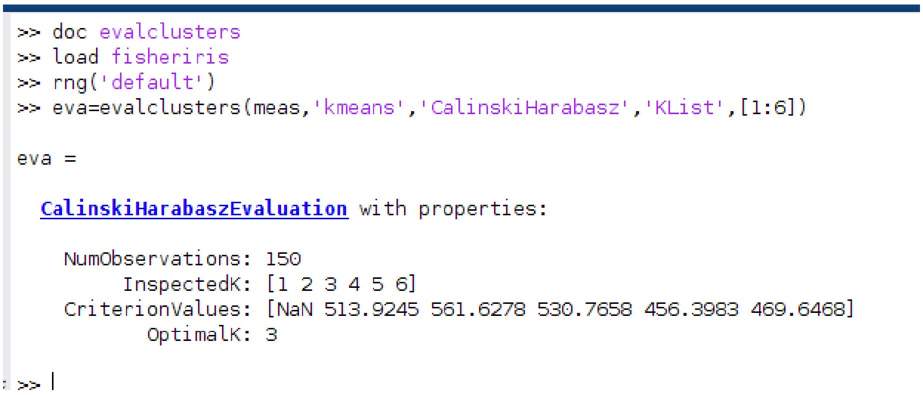
(b)



3.2 k-means initialization

The first centroid is 6,the second centroid is 33 and the third centroid is 20.Then using $k-$means algorithm ,we can get three clusters as $\{6, 8, 10\}, \{20, 26\}, \{30, 32, 33\}$

3.3 Determine number of clusters



```
>> doc evalclusters
>> load fisheriris
>> rng('default')
>> eva=evalclusters(meas,'kmeans','CalinskiHarabasz','KList',[1:6])

eva =

  CalinskiHarabaszEvaluation with properties:

    NumObservations: 150
         InspectedK: [1 2 3 4 5 6]
    CriterionValues: [NaN 513.9245 561.6278 530.7658 456.3983 469.6468]
           OptimalK: 3

>> |
```