POORNIMA

LECTURE NOTES

Campus: PCE Course: BTECH Class Section: IIIrd CSE-A 24-02-21

Name of Faculty: Praveen Kr. Gaur Name of Subject: Machine Learning 06

Date (Prep): 24-02-21 Date (DoD): 5-02-21 Unit No/Topic: 6

OBJECTIVE: To be written before taking the lecture (Pl. write in bullet points the main topics you are going to discuss in the lecture)

Decision Tree and Random Forest.

IMPORTANT & RELEVANT QUESTIONS:

what is entropy?

FEED BACK QUESTIONS (AFTER 20 MINUTES):

How we determine the best spilt for a decision Tree.

OUTCOME OF THE DELIVERED LECTURE: To be written after taking the lecture (Pl. write in bullet points about students' feedback on this lecture, level of understanding of this lecture by students etc.)

good

REFERENCES: Text/Ref. Book with Page No. and relevant Internet Websites:

scikit with learn ebook

Decision Tree - Important Terms -

① entropy - is the measure of randomness or predictability in the dataset.

eg.

② Information gain - It is the measure of decrease in entropy after the dataset is split.

③ Leaf-NODE - Leaf node carries the classification or the decision.

④ Root NODE - Top most node.

We have to frame the condition in such a way that the information gain is the highest.

$$\sum_{i=1}^{n} P(value\,i) \cdot \log_2(P(value\,i))$$

$$\frac{3}{8}\log_2\frac{3}{8} + \frac{2}{8}\log\frac{2}{8} + \frac{1}{8}\log_2\frac{1}{8} + \frac{2}{8}\log_2\left(\frac{2}{8}\right)$$

Entropy = 0.571

PAGE NO. .....

Random Forest Algo^m

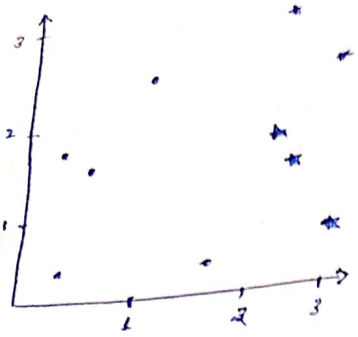
No overfitting - Use of multiple trees, reduces the risk of overfitting.

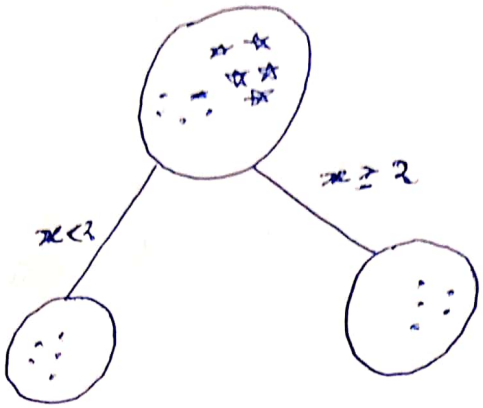
High accuracy - Runs efficiently on large database highly accurate predictions

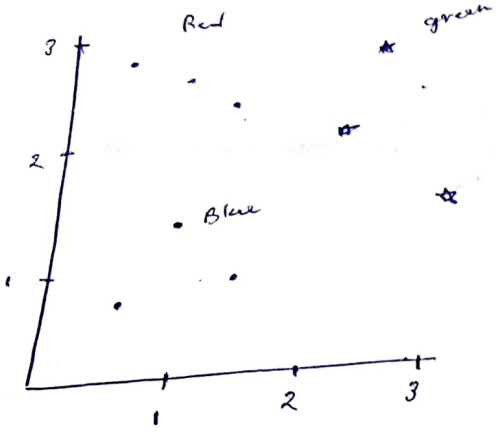
Estimates missing data - RF can maintain accuracy when a large proportion of data is missing.

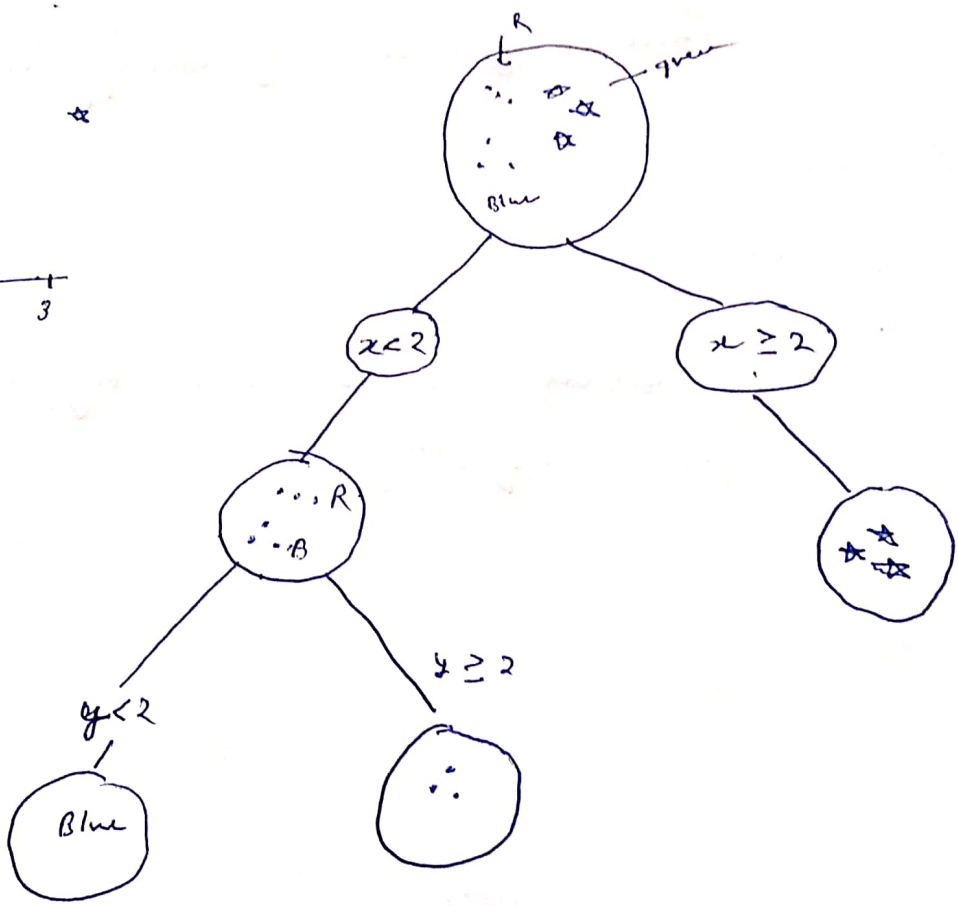
RF or R Decision Forest - is a method that operates by constructing multiple decision trees during training phase.

↳ The decision of the majority of the trees is chosen by the random forest as final decision.









Red.

PAGE NO.

Decision Tree in Machine Learning:-

(Set of nested - if and if else if conditions)

what is Decision Tree :- Decision tree is a tree shaped diagram used to determine a course of action. Each branch of the tree represents a possible action/decision occurence or reaction.

↳ it uses a tree like model to represent decision.

↳ Decision Tree can be used to solve the classification problems. (For eg- Discriminating the certain type of vegtables based on certain type of features.)

↳ set of axis hyperparallel planes are as axis parallel in a DT.

Advantages- 1. its simple to understand, interpret and visualize.

2. Little effort required for data preparation.

3. you can handle both numerical and categorical data.

4. Non - linear paramder don't effect its performance.

Disadvantages of Decision Tree:-

1). overfitting occurs when the algorithm captures noise in the data.

2). The modal can get unstable due to small variation in data (low biased tree).

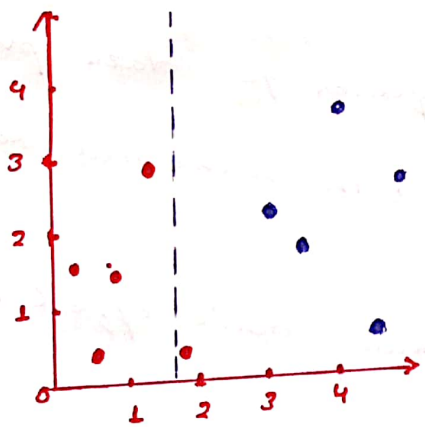
3). A highly complicated decision tree tends to have a low bias which makes difficult for the modal to work with new data.

Decision Tree - (Important Terms)

1. Entropy:- Entropy is the measure of Randomness or unpredictability in the dataset.

Initially the dataset is having higher entropy).

Entropy is a metric used to train decision tree. These metric measures the "quality of a split".



Left branch has 4 red dots.

Right branch has 1 red and 5 blue dots

(A data set having only blues would have very low or zero entropy).

(A dataset having mixed blues, greens, and red would have relatively high entropy.)

Information Entropy for a dataset with c classes are denoted as:-

$$E = -\sum_{i}^{K} P_i \log_2 P_i$$

(The no of data classes / or c)

where P_i = is the probability of randomly picking an i^{th} element of class c.

For eg- consider a dataset with 1 blue, 2 black, 3 reds. (• •• •••)

Then $E = -(P_b \log_2 P_b + P_{black} \log_2 P_{black} + P_R \log_2 P_R)$

Here $P_b = 1/6$ $\quad P_{black} = 2/6$ $\quad P_R = 3/6$

So $E = -(1/6 \log_2 1/6) + 2/6 \log_2 (2/6) + 3/6 \log_2 (3/6))$

$= 1.46$

Eg-2. Consider 3 blues ... Entropy would be

$E = -(1 (\log 1) = 0$

So before spilt - we have 5 blues and 5 red. So Entropy was $E_{before} = -(0.5 \log_2 0.5 + 0.5 \log_2 0.5)$

$= 1$

After spilt, we have two branches - Left branch - 4 reds and Right branch have 1 red and 5 blues.

So Entropy after spilt.

$E_{left} = -(0.8 \log 1) = 0$

$E_{right} = -(1/6 \log_2 (1/6)) + 5/6 \log_2 5/6))$

$= \boxed{0.65}$

$E_{split} = 0.4 * 0 + 0.6 * 0.65$

$= 0.39$

we started with $E_{before} = 1$ entropy before the split and now are down to 0.39.

So Information Gain = $1 - 0.39 = 0.61$

higher Information Gain = More Entropy removed.

Information Gain:- It is the measure of decrease in entropy after the dataset is split. Information gain is calculated for a split by subtracting weighted entropies of each branch from the original entropy.

LEAF NODE:- Leaf node carries the classification.

ROOT NODE:- The top most decision node is known as the Root node.

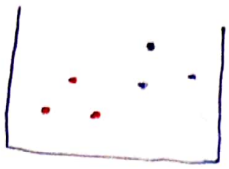
Note- we have to frame the conditions such that it split the data in such way that the information gain is the highest.

when training a decision tree using these metrics, the best split is chosen by maximizing the Information Gain.

Build a Decision Tree

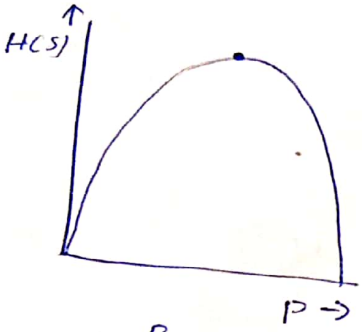
$$H(S) = -\sum_{i=1}^{c} P_c \log P_c$$

P_c is Prob. of class c.



$$H(S) = -\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{2}\log\tfrac{1}{2}\right)$$
$$= 1$$

$$H(S) = 1 \log 1$$
$$= 0$$



P = 0.5 highest entropy.

classes [yes → 9, No → 5]

$$H(S) = -\left(\tfrac{9}{14}\log\tfrac{9}{14} + \tfrac{5}{14}\log\tfrac{5}{14}\right)$$
$$= 0.41 + 0.53$$
$$H(S) = 0.94$$

(a H)

$$IG(S, A) = H(S) - \sum \frac{|S_v|}{|S|} H(S_v)$$

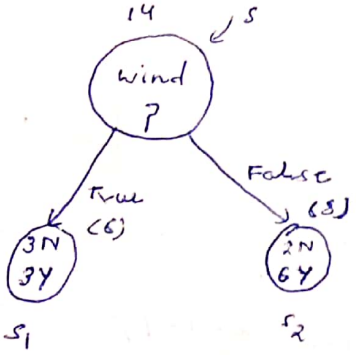
$H(S)$ ↓ (old entropy)

$\frac{|S_v|}{|S|}$ ↑ Total eg

New entropy

New set

(Ratio of no of eg in new set over all sets we have)



Maximize the information gain $H(S) - \sum \frac{5}{14}\left[-\frac{6}{8}\log\frac{6}{8} + \frac{2}{8}\log\frac{2}{8}\right]$

$$+ \frac{6}{14}\left[-\frac{3}{6}\log\frac{3}{6} - \frac{3}{6}\log\frac{3}{6}\right]$$

$$= H(S) - \tfrac{8}{14}(0.81) + \tfrac{6}{14}$$
$$= 0.94 - 0.892$$
$$= 0.0048 \leftarrow$$

IG(S, wind) = 0.048

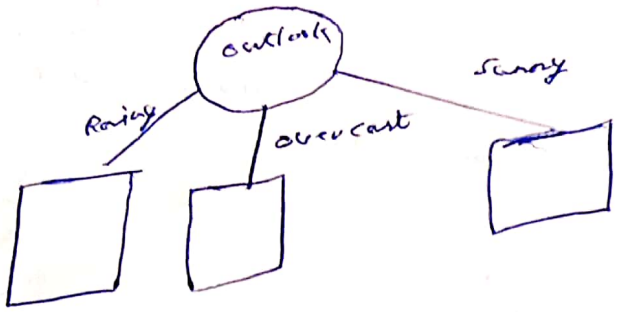
IG(S, outlook) = 0.247 ✓

IG(S, Humidity) = 0.15

IG(S, Temp) = 0.029

Do yourself.

So Here we choose IG(S, outlook)



Repeat this process for other attributes recursively.

until entropy is zero..

=