4. decision tree

~~classification :-~~

10] What is entropy? How is it calculated?

→ ① entropy is the measures of impurity, disorder or uncertainly in bunch of examples.

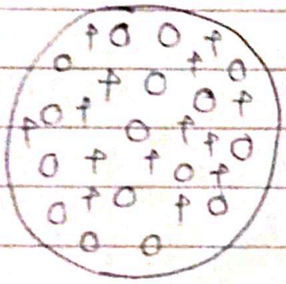
② entropy = $\sum_i -p_i \log_2 p_i$

p_i = probability of class i

③ The higher the entropy more the information content.

④ entropy should be always close to 1.

ex.



16/30 are circles :-

$\log_2^{(16/30)} = -0.009$

$= -0.9$

14/30 are crosses : $- -0.011$

$= -1.1$

entropy $= -\left(\frac{16}{30}\right)(-0.9) - \left(\frac{14}{30}\right)(-1.1)$

$= 0.99$

x—o—x

entropy
Gini index
Info gain
ID3

Q.7 What are merits & demerits of decision tree?

→ **Merits of decision tree :-**

① easy to understand :-
- Decision tree output is very easy to understand
- It does not requiee any stastical knowledge
- Its graphical representation is very inherit and users can easily relate their hypothesis

② Useful in Data exploration :-
- D.T. is one of fastest way to identify most significant variable & relation betn two or more variables.
- With help of DT we can create new variables.

③ Less Data cleaning required :-
- It requires less data cleaning compared to some other modelling techniques.

④ Data type is not a constraint :-
- It can handle both numerical & categorical variable.

⑤ Non parametric method :-
- DT have no assumption about space distribution & classifier structure.

Demerits of DT :-

① Over fitting :-
- It is the most practical difficulty for DT models.
- This problem gets solved by setting constraint on model parameters & pruning

② Not fit for continuous variable :-
— While working with continuous numerical variable, DT looses info.

Q·17] What is gini index? How is it calculated?

→ ① It $Gini = 1 - \sum_{i=1}^{k} P_k^2$

where P_k denotes the proportion of instances belonging to class k $(k = 1, ---, k)$

P_k = probability of class k

② It is calculated by subtracting the sum of squared probabilities of each class from one.

③ The classic CART alg. uses the Gini index for constructing DT.

④ Gini index of pure / homogeneous data is 0.

⑤ Gini index of impure data greater that 0.

ex

$D = \{Y, Y, Y, Y, N, N, N\}$

Total element = 7

$Y = 4 \quad N = 3$

$Gini(D) = \quad Y = 1 - (4/7)^2 - (3/7)^2$

$= 1 - 0.3265 - 0.1837$

$= 0.4898$

Unit-4. Decision Tree

* What is Decision Tree?

① Decision tree is a type of supervised Learning algorithm mostly used for classification problem.

② The tree consists of decision nodes & leaf nodes.

③ The decision node has two or more branches each representing values for the attribute tested.

④ A leaf node attribute produces a homogeneous result, which does not require additional classification testing.

* characteristics :-

① every non leaf node (decision node) represents an attribute in dataset.

② every branch represents possible values of an attribute.

③ Every leaf node represent the value of target attribute.

④ starting node is called as root node.

⑤ To make a decision, the flow starts at root node, navigates through the arc/edges until it reaches a leaf node, & then makes decision Based on leaf node value.

* How to build decision tree?

1. ID3 (Iterative Dichotomiser3) → uses entropy funⁿ & information gain as metrics.
2. CART (classification & Regression Trees) → uses Gine Index (classificatⁿ) as metric.

ex. 2] for set R = {a, a, a, b, b, b, b, b}

entropy (R) = I(R) = $-\left[\left(\frac{3}{8}\right)\log_2\left(\frac{3}{8}\right) + \left(\frac{5}{8}\right)\log_2\left[\frac{5}{8}\right]\right]$

a-val b-val

* entropy should be always close to 1. *

* Information Gain :-

① Information gain tell us how important a given attribut is from all other attributes.
② The information gain is the reduction in 'uncertainty' when choosing an attribute.

x-o-x

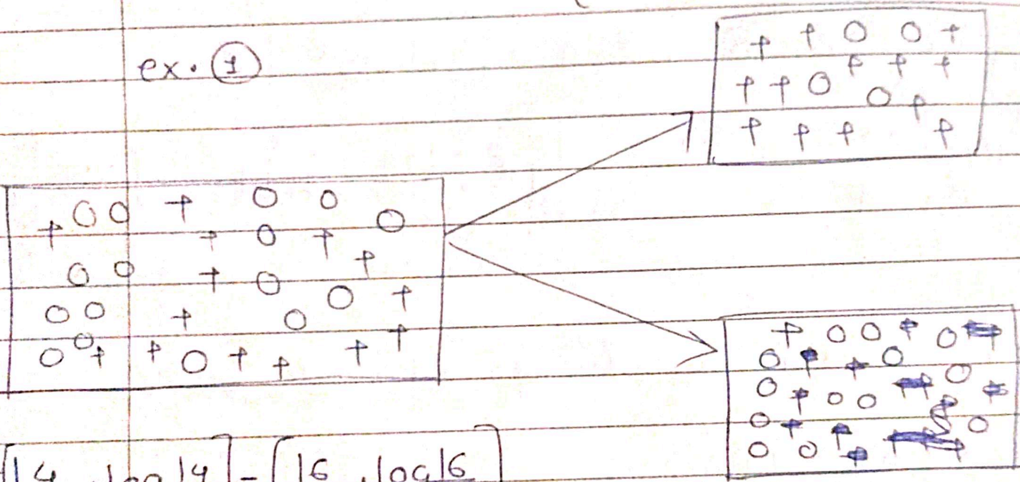
① The information gain is based on the decrease in entropy after a dataset is split on an attribute.
② The attribute that yields the largest IG is chosen for decision node.
③ A branch set with entropy of 0 is a leaf node.

Information Gain (IG) = entropy − average entropy

* All logs are with respect to base 2.

$-\left[\frac{13}{17}\cdot\log_2\frac{13}{17}\right]-\left[\frac{4}{17}\cdot\log_2\frac{4}{17}\right] = 0.72$

ex. 1



$-\left[\frac{14}{30}\cdot\log_2\frac{14}{30}\right]-\left[\frac{16}{30}\cdot\log_2\frac{16}{30}\right]$

$= 0.996$

$-\left[\frac{1}{13}\cdot\log_2\frac{1}{13}\right]-\left[\frac{12}{13}\cdot\log_2\frac{12}{13}\right] = 0.391$

$IG = 0.996 - 0.615 = \underline{0.38}$

entropy & information gain: -

- Advantages of ID3 :-

1. Understandable prediction ~~values~~ rules are created from the training data.
2. build the fastest tree.
3. build short tree.
4. Only need to test enough attributes untill all data is classified.
5. whole dataset is searched to create tree.

- Disadv. :-

1. Data may be over-fitted or over-classified, if a small sample is tested.
2. Only one attribute at a time is tested for making a decision.