DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name: Gadilohar Hitesh Digambar
Expt. Title: Implement the find-s Inductive learning algm
Class: F.Y. MCA   Batch: ____   Performed on: ____
Roll No. 38   Expt. No. 01   Submitted on: ____
Remarks: ____   Returned on: ____

* Find S Algorithm :-

The find s algorithm is a basic Concept learning algorithm in maction learning. The find-s algorithm finds the most Specific hypothesis that fits all the positive examples. We have to note hare that the algorithm Consider only those positive traning example.

Algorithm :-

Step1 :- Initialize h to the most Specific hypothesis in H

Step2 :- for each positive trining instance x
for each attribute Constraint a_i in h If the Constraint a is Satisfied by x Then do nothing.
Else replace a in h by the next more general Constraint that is Satisfied by x

Step 3 :- Output hypothesis h.

Example	Color	Toughness	Fungus	Appearance	Poisonous
1	Green	Hard	No	Wrinkled	Yes
2	Green	Hard	Yes	Smoth	No
3	Brown	Soft	No	Wrinkled	No
4	Orange	Hard	No	Wrinkled	Yes
5	Green	Soft	Yes	Smooth	Yes

Complete for :
Algorithm
Chart
Programme Listing
Results
Comments

frist we Consider the hypothesis to be a more Specific hypothesis Hence our hypothesis woul

$h = \{ \phi, \phi, \phi, \phi, \phi, \phi, \}$

Consider example 1 :-

h = { Green, Hard, No, Wrinkled }

Consider example 2 :-

h = { Green, Hard, No, Wrinkled }

Consider example 3 :-

h = { Green, Hard, No, wrinkled }

Have we See that above 3 examples has a negitive Outcome Hence we neglect this example & our hypothesis remains the Same.

Consider Example 4 :-

h = { ?, Hard, No, Wrinkled }

Consider Example 5 :-

h = { ?, ?, ?, ? }.

Hence for the Givan data find final hypothesis would be:

Final Hypothesis : h = { ?, ?, ?, ? }

Scanned with CamScanner

DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name Gadilohar Hitesh Digambar
Expt. Title Implement the Candidate-Elimination Inductive learning algorithm
Class F.Y. MCA Batch Performed on
Roll No. 58 Expt. No. 02 Submitted on
Remarks Returned on

* Candidate - Elimation Algorithm :-

The Candidate elimation algorithm incrementally builds the versition Spece given a hypothesis Space H & a Set E of examples.

Algorithm :-

Step 1 :- Load Data Set.

Step 2 :- Initialize General Hypothesis & Specific Hypothesis.

Step 3 :- for each training example.

Step 4 :- If example is positive example.

if attribute_value == hypothesis_value :

Do nothing.

else :

replace attribute value with '?'

(Basically generalizing it)

Step 5 :- if example is negetive example make genralizing hypothesis more Specific.

Example :-

Consider the dataset given below:

te for :
hm
hart
mme Listing
s
ents

Scanned with CamScanner

Sky	Tempodure	Humid	wind	Water	forest	Output
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	high	strong	Warm	Same	Yes
rainy	Cold	high	strong	Warm	Chnge	no
Sunny	Warm	high	Strong	Cool	Chnge	Yes

Initially :-

G = [[?, ?, ?, ?, ?, ?], [?, ?, ?, ?, ?, ?], [?, ?, ?, ?, ?,
[?, ?, ?, ?, ?, ?], [?, ?, ?, ?, ?, ?] [?, ?, ?

S = [Null, Null, Null, Null, Null, Null]

for instance 1:- <'Sunny', 'Worm', 'Normal', 'Strong', '
& positive Output.

G1 = G

S1 = ['sunny', 'warm', 'normal', 'strong', 'warm', 'Same

for instance 2:-

<'Sunny', 'warm', ?, 'Strong', 'Warm', 'Same']

for instance 3:-

<'rainy', 'Cold', 'high', 'Strong', 'Warm', 'Chnge'> &

G3 = [['Sunny', ?, ?, ?, ?, ?], [?, 'warm', ?, ?, ?, ?], [?, ?

Output [?, ?, ?, ?, ?, ?] [?, ?, ?, ?, ?, ?] [?, ?, ?, ?, ?, 'Sam

S3 = S2

for instance 4:- <'Sunny', 'Warm', 'high', 'Strong', 'cool',

G4 = G3

S4 = ['Sunny', 'Warm', ?, 'Strong', ?, ?]

At last, by Synchronizing the G4 & S5 algorithm

Output:-

G = [['Sunny', ?, ?, ?, ?, ?], [?, 'Warm', ?, ?, ?, ?]]

S = ['Sunny', 'Warm', ?, 'Strong', ?, ?]

DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name Gadilohar Hitesh Digambar
Expt. Title Write a program to implement Design tree using python programming of your own.
Class FY MCA Batch ______ Performed on ______
Roll No. 38 Expt. No. 03 Submitted on ______
Remarks ______ Returned on ______

* What is decision trees?
Decision tree is a Supervised learning technique that can be used for both classification & regression problems, but mostly it is preferred for solving classification program. It is a tree structure classifier where internal nodes represent the feature of a dataset. branches represent the design rules & each leaf node represent the outcome.
It is a graphical represent of for getting all the possible solution to problem/decision based on given conditions. this works surprisingly well in most of class.

* ID3 Algorithm:-
- If all examples have same lable:-
  - return a leaf with that lable
- Else if their are no feature leaf to test:-
  - return a leaf with the most common lable
- Else:-
  - chose the feature F that maximise the information gain of S to be the next node using equation(2.3)
  - Add branch form the node for each possible value F in $\hat{F}$.

- For each Branch:-
  * Calculate S_f by removing $\hat{f}$ from the set of
  * recursively call the algorithm with S_f to comp the gain relative to the current set of exa

* formulla of Information Gain :.

1) Entropy :- $-P \log_2 p - q \log_2 q$.

- To build a decision tree, we can need to calculate two types of entropy using frequency tables:

i) Entropy using the frequency table of one

$$E(S) = \sum_{i=1}^{c} -P_i \log_2 P_i$$

ii) Entropy using the frequency table of two attr

$$E(T, X) = \sum_{c \in X} P(c) \cdot E(c)$$

2) $Gini = 1 - \sum_{i=1}^{n} (P_i)^2$

3) Information Gain $= 1 -$ Entropy.

Scanned with CamScanner

DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name Gadilohar Hitesh Digambar
Expt. Title White Program to Calculate popular Attribute.
Class MCA I Batch B1 Performed on
Roll No. 38 Expt. No. 04 Submitted on
Remarks Returned on

Attribute Selection Measure: (ASM)

The best attribute or feature is Selected using the attribute Selection measure (ASM) the attribute. Selection is the root node feature.

Asm is a technique used for the Selecting best attribute for discrimation among tuples it give rank to each attribute & the best attribute is Selected as spliting criterion.

The most popular methods of Selection are:-

1) Information Gain:-

Information gain is a decision in decrese in entropy Descision tree none use of Information gain and entropy to determine which feature to Split into nodes to get closer to predicting the target & also to determine when to stop spliting.

Information Gain = Entropy (s) - [weighted Avg) * Entropy (each features).

2) Gini Index:-

Gini index is a measure of impurity or purity used while Creating a decision tree in the CART index Can be Calculate using the bellow formulla

Gini index = $1-\sum_i p_i^2$

complete for:
Algorithm
Flow Chart
Programme Listing
Results
Comments

DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name Gadilohar Hitesh Digambar
Expt. Title
Class F.Y. MCA Batch Performed on
Roll No. 38 Expt. No. 05 Submitted on
Remarks Returned on

* K-Nearest Neighbour (KNN) Algorithm:-

K-Nearest Neighbour is one of the simple Machine learing algorithm based on Supervised lerny technique ~~K~~ NN algorithm Store all the avalable data & classification a new data point based on the Similarity. This mean when new data appears. then it can be easlly classified into a well suit Category by using K-NN algorithm.

K-NN is a non-parametric algorithm which mean it does not make any assumption on underlying data. It is a lazy learning algorithm where all Computation is defered until classification.

* Algorithm:-

- The KNN Classification is a performed using the followng four steps
- Compute the distance metric betn the test data point & all the labeled data point.
- Order the lable data point in the incresing order of the distonce metric.
- Select the top k labled data point & look at the class lables.
- find the class lable that the majority of these lable data points have & assign in to the test data point.

* Distance Calculation formulla :-

1) Eucliden Distance :-

It is generally used to find the distance bet

real-valued vectors.

$$d = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

2) Manhatten Distance :-

This is simple way or technique to calcul

distance bet^n two point of ten called Texicab dis

or city block distance.

Manhatten Distence = Sum for i to N

Sum $||x_{1i} - x_{2i}||$

3) Hamming distance :-

The haming distence is mostly used

processing or having the boolean Vector Boolean

Mean the data is in the form of binary d

1.

Hamming distence (x1, x2) = ...

4) Minkowski Distance :-

Minkowski distence is a generation Org generdl

form of the eucidean & manhatten distanc

$$||x_1 - x_2|| = \left(\sum_{i=1} |x_{1i} - x_{2i}|^p\right)^{1/p}$$