DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name Toke Pratiksha Raju
Expt. Title Write pgm to calculate popular attribute selection measures (ASM) like
Class FYMCA Batch B-4 Performed on
Roll No. 130 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 selected is the root node feature.

ASM is a technique used for selecting best attribute for discrimination among tuple It gives rank to each attribute and the best attribute is selected by spliting creation.

The most popular method of selection are:

1) Information Gain:

Information gain is decrease in entropy Decision trees makes use of information gain and entropy to determine which feature to split into node to get closer to predicting target and also to determine when to stop spliting

Information Gain = Entropy (s) - [(weighted Avg) * Entropy (each feature)]

2] Gini Index

Gini Index is measure of impurity or purity used while creating a decision tree in CART (classification & regression Tree algorithm gini index can be calculated by using below formula

Gini Index = $1 - \Sigma_j p_j^2$

Scanned with CamScanner

3] Gain Ratio:-

The gain ratio is the modification of information gain. It takes into account number and size of branches when choosing an attribute

It takes intrinsic information into account

GR(S.A) = Gain(S.A) / Int I(S.A)