Assignment 2: Report

ID3 Algorithm Implementation

Pushpita Panigrahi pxp171530

Sidharth Swarup Panda ssp171730

# Accomplishments and Learnings

## The ID3 algorithm does binary classification of the data provided to it and predicts the output class(binary) of a given instance and thus finally calculating the accuracy over the entire dataset.

We learnt the supervised machine learning technique and understood how to train and test the machine when given a set of data. We also learned that as we go on creating deeper trees, at first the accuracy may look perfect but on testing data the the accuracy may become lower. This is a case of overfitting which we can avoid by pruning the tree so that there is predictability and not rote learning of the machine. We created and traversed a tree by using DFS and assigning class labels as number of positive and negative instances that are held by each node. By convention we assigned 0 to left side and 1 to right side of the tree and 0 to denote negative instances and 1 to denote positive instances.

It was a great exercise to understand how ID3 algorithm runs in the background.

We got best accuracy of 100% on training and 75% on test before pruning and 96.5% on training and 75.5% on test after pruning on Dataset1.

## Assumptions

1. We find the set of nodes to be pruned at once. If while pruning a node and its sub tree, other nodes in the chosen set are deleted, we ignore those nodes.
2. We assume that the datasets have a column name “Class” which contains the classification types.
3. We assume that all the attributes and class values are binary.
4. We do pruning from level 3 of the tree as it does not make sense to delete the root node or the first level would prune a very large part of the tree.
5. If in 3 iterations of pruning we do not get an increased accuracy, we stop there and do not go for further pruning.

## Best Results

Command : python id3.py data\_sets1/training\_set.csv data\_sets1/validation\_set.csv data\_sets1/test\_set.csv 0.06

Output Screenshot :

XO=0:

|  XM=0:

|  |  XF=0:

|  |  |  XB=0:

|  |  |  |  XG=0: 0

|  |  |  |  XG=1:

|  |  |  |  |  XD=0:

|  |  |  |  |  |  XS=0: 0

|  |  |  |  |  |  XS=1:

|  |  |  |  |  |  |  XC=0: 1

|  |  |  |  |  |  |  XC=1:

|  |  |  |  |  |  |  |  XH=0: 0

|  |  |  |  |  |  |  |  XH=1: 1

|  |  |  |  |  XD=1:

|  |  |  |  |  |  XP=0:

|  |  |  |  |  |  |  XE=0: 0

|  |  |  |  |  |  |  XE=1: 1

|  |  |  |  |  |  XP=1: 0

|  |  |  XB=1:

|  |  |  |  XD=0: 0

|  |  |  |  XD=1:

|  |  |  |  |  XI=0: 0

|  |  |  |  |  XI=1:

|  |  |  |  |  |  XN=0: 0

|  |  |  |  |  |  XN=1: 1

|  |  XF=1: 0

|  XM=1:

|  |  XB=0:

|  |  |  XF=0:

|  |  |  |  XG=0: 0

|  |  |  |  XG=1:

|  |  |  |  |  XD=0: 1

|  |  |  |  |  XD=1:

|  |  |  |  |  |  XC=0:

|  |  |  |  |  |  |  XS=0: 0

|  |  |  |  |  |  |  XS=1:

|  |  |  |  |  |  |  |  XR=0: 1

|  |  |  |  |  |  |  |  XR=1: 0

|  |  |  |  |  |  XC=1: 0

|  |  |  XF=1:

|  |  |  |  XS=0:

|  |  |  |  |  XQ=0:

|  |  |  |  |  |  XI=0: 0

|  |  |  |  |  |  XI=1:

|  |  |  |  |  |  |  XP=0: 0

|  |  |  |  |  |  |  XP=1: 1

|  |  |  |  |  XQ=1:

|  |  |  |  |  |  XI=0: 1

|  |  |  |  |  |  XI=1:

|  |  |  |  |  |  |  XL=0: 1

|  |  |  |  |  |  |  XL=1:

|  |  |  |  |  |  |  |  XT=0: 0

|  |  |  |  |  |  |  |  XT=1: 1

|  |  |  |  XS=1:

|  |  |  |  |  XR=0:

|  |  |  |  |  |  XD=0: 1

|  |  |  |  |  |  XD=1:

|  |  |  |  |  |  |  XP=0:

|  |  |  |  |  |  |  |  XT=0:

|  |  |  |  |  |  |  |  |  XQ=0: 0

|  |  |  |  |  |  |  |  |  XQ=1: 1

|  |  |  |  |  |  |  |  XT=1: 1

|  |  |  |  |  |  |  XP=1: 0

|  |  |  |  |  XR=1:

|  |  |  |  |  |  XI=0:

|  |  |  |  |  |  |  XU=0: 1

|  |  |  |  |  |  |  XU=1: 0

|  |  |  |  |  |  XI=1:

|  |  |  |  |  |  |  XG=0:

|  |  |  |  |  |  |  |  XQ=0:

|  |  |  |  |  |  |  |  |  XJ=0:

|  |  |  |  |  |  |  |  |  |  XU=0: 1

|  |  |  |  |  |  |  |  |  |  XU=1: 0

|  |  |  |  |  |  |  |  |  XJ=1: 1

|  |  |  |  |  |  |  |  XQ=1: 0

|  |  |  |  |  |  |  XG=1: 1

|  |  XB=1:

|  |  |  XI=0: 0

|  |  |  XI=1:

|  |  |  |  XK=0:

|  |  |  |  |  XD=0: 0

|  |  |  |  |  XD=1:

|  |  |  |  |  |  XP=0: 1

|  |  |  |  |  |  XP=1:

|  |  |  |  |  |  |  XS=0:

|  |  |  |  |  |  |  |  XG=0: 1

|  |  |  |  |  |  |  |  XG=1:

|  |  |  |  |  |  |  |  |  XN=0: 0

|  |  |  |  |  |  |  |  |  XN=1: 1

|  |  |  |  |  |  |  XS=1: 0

|  |  |  |  XK=1: 0

XO=1:

|  XI=0:

|  |  XM=0:

|  |  |  XQ=0:

|  |  |  |  XF=0:

|  |  |  |  |  XH=0:

|  |  |  |  |  |  XS=0: 0

|  |  |  |  |  |  XS=1:

|  |  |  |  |  |  |  XP=0: 0

|  |  |  |  |  |  |  XP=1: 1

|  |  |  |  |  XH=1: 1

|  |  |  |  XF=1: 0

|  |  |  XQ=1:

|  |  |  |  XJ=0:

|  |  |  |  |  XN=0:

|  |  |  |  |  |  XP=0: 1

|  |  |  |  |  |  XP=1:

|  |  |  |  |  |  |  XU=0: 0

|  |  |  |  |  |  |  XU=1:

|  |  |  |  |  |  |  |  XF=0: 0

|  |  |  |  |  |  |  |  XF=1: 1

|  |  |  |  |  XN=1: 0

|  |  |  |  XJ=1:

|  |  |  |  |  XL=0:

|  |  |  |  |  |  XH=0: 0

|  |  |  |  |  |  XH=1:

|  |  |  |  |  |  |  XU=0: 1

|  |  |  |  |  |  |  XU=1:

|  |  |  |  |  |  |  |  XK=0: 0

|  |  |  |  |  |  |  |  XK=1: 1

|  |  |  |  |  XL=1: 1

|  |  XM=1:

|  |  |  XQ=0:

|  |  |  |  XF=0:

|  |  |  |  |  XL=0:

|  |  |  |  |  |  XC=0: 1

|  |  |  |  |  |  XC=1:

|  |  |  |  |  |  |  XH=0: 1

|  |  |  |  |  |  |  XH=1:

|  |  |  |  |  |  |  |  XU=0:

|  |  |  |  |  |  |  |  |  XR=0:

|  |  |  |  |  |  |  |  |  |  XT=0: 0

|  |  |  |  |  |  |  |  |  |  XT=1: 1

|  |  |  |  |  |  |  |  |  XR=1: 0

|  |  |  |  |  |  |  |  XU=1: 1

|  |  |  |  |  XL=1:

|  |  |  |  |  |  XP=0: 0

|  |  |  |  |  |  XP=1:

|  |  |  |  |  |  |  XT=0: 1

|  |  |  |  |  |  |  XT=1:

|  |  |  |  |  |  |  |  XJ=0: 1

|  |  |  |  |  |  |  |  XJ=1: 0

|  |  |  |  XF=1: 0

|  |  |  XQ=1: 0

|  XI=1:

|  |  XT=0:

|  |  |  XH=0:

|  |  |  |  XP=0:

|  |  |  |  |  XF=0: 0

|  |  |  |  |  XF=1:

|  |  |  |  |  |  XR=0:

|  |  |  |  |  |  |  XL=0:

|  |  |  |  |  |  |  |  XU=0: 0

|  |  |  |  |  |  |  |  XU=1: 1

|  |  |  |  |  |  |  XL=1: 1

|  |  |  |  |  |  XR=1:

|  |  |  |  |  |  |  XQ=0: 1

|  |  |  |  |  |  |  XQ=1: 0

|  |  |  |  XP=1:

|  |  |  |  |  XS=0:

|  |  |  |  |  |  XD=0:

|  |  |  |  |  |  |  XC=0:

|  |  |  |  |  |  |  |  XU=0:

|  |  |  |  |  |  |  |  |  XQ=0: 1

|  |  |  |  |  |  |  |  |  XQ=1: 0

|  |  |  |  |  |  |  |  XU=1:

|  |  |  |  |  |  |  |  |  XG=0: 1

|  |  |  |  |  |  |  |  |  XG=1: 0

|  |  |  |  |  |  |  XC=1: 0

|  |  |  |  |  |  XD=1:

|  |  |  |  |  |  |  XM=0:

|  |  |  |  |  |  |  |  XJ=0: 0

|  |  |  |  |  |  |  |  XJ=1: 1

|  |  |  |  |  |  |  XM=1: 1

|  |  |  |  |  XS=1: 1

|  |  |  XH=1:

|  |  |  |  XJ=0:

|  |  |  |  |  XC=0:

|  |  |  |  |  |  XN=0: 1

|  |  |  |  |  |  XN=1:

|  |  |  |  |  |  |  XG=0:

|  |  |  |  |  |  |  |  XF=0: 1

|  |  |  |  |  |  |  |  XF=1: 0

|  |  |  |  |  |  |  XG=1: 0

|  |  |  |  |  XC=1:

|  |  |  |  |  |  XM=0: 0

|  |  |  |  |  |  XM=1:

|  |  |  |  |  |  |  XS=0:

|  |  |  |  |  |  |  |  XR=0: 1

|  |  |  |  |  |  |  |  XR=1: 0

|  |  |  |  |  |  |  XS=1: 1

|  |  |  |  XJ=1:

|  |  |  |  |  XS=0: 1

|  |  |  |  |  XS=1:

|  |  |  |  |  |  XG=0:

|  |  |  |  |  |  |  XB=0: 0

|  |  |  |  |  |  |  XB=1:

|  |  |  |  |  |  |  |  XN=0: 1

|  |  |  |  |  |  |  |  XN=1:

|  |  |  |  |  |  |  |  |  XU=0: 1

|  |  |  |  |  |  |  |  |  XU=1: 0

|  |  |  |  |  |  XG=1:

|  |  |  |  |  |  |  XQ=0: 1

|  |  |  |  |  |  |  XQ=1:

|  |  |  |  |  |  |  |  XU=0: 1

|  |  |  |  |  |  |  |  XU=1: 0

|  |  XT=1:

|  |  |  XS=0:

|  |  |  |  XQ=0:

|  |  |  |  |  XK=0:

|  |  |  |  |  |  XC=0:

|  |  |  |  |  |  |  XR=0:

|  |  |  |  |  |  |  |  XH=0:

|  |  |  |  |  |  |  |  |  XE=0: 0

|  |  |  |  |  |  |  |  |  XE=1: 1

|  |  |  |  |  |  |  |  XH=1: 1

|  |  |  |  |  |  |  XR=1:

|  |  |  |  |  |  |  |  XU=0: 0

|  |  |  |  |  |  |  |  XU=1:

|  |  |  |  |  |  |  |  |  XL=0: 0

|  |  |  |  |  |  |  |  |  XL=1: 1

|  |  |  |  |  |  XC=1: 1

|  |  |  |  |  XK=1:

|  |  |  |  |  |  XD=0:

|  |  |  |  |  |  |  XJ=0: 0

|  |  |  |  |  |  |  XJ=1: 1

|  |  |  |  |  |  XD=1: 0

|  |  |  |  XQ=1:

|  |  |  |  |  XM=0:

|  |  |  |  |  |  XN=0:

|  |  |  |  |  |  |  XU=0: 1

|  |  |  |  |  |  |  XU=1: 0

|  |  |  |  |  |  XN=1:

|  |  |  |  |  |  |  XR=0: 1

|  |  |  |  |  |  |  XR=1:

|  |  |  |  |  |  |  |  XL=0: 1

|  |  |  |  |  |  |  |  XL=1: 0

|  |  |  |  |  XM=1: 0

|  |  |  XS=1:

|  |  |  |  XL=0:

|  |  |  |  |  XD=0:

|  |  |  |  |  |  XU=0: 1

|  |  |  |  |  |  XU=1:

|  |  |  |  |  |  |  XQ=0:

|  |  |  |  |  |  |  |  XG=0: 1

|  |  |  |  |  |  |  |  XG=1: 0

|  |  |  |  |  |  |  XQ=1:

|  |  |  |  |  |  |  |  XH=0: 0

|  |  |  |  |  |  |  |  XH=1: 1

|  |  |  |  |  XD=1:

|  |  |  |  |  |  XG=0: 0

|  |  |  |  |  |  XG=1: 1

|  |  |  |  XL=1:

|  |  |  |  |  XH=0:

|  |  |  |  |  |  XQ=0:

|  |  |  |  |  |  |  XD=0: 0

|  |  |  |  |  |  |  XD=1:

|  |  |  |  |  |  |  |  XR=0: 0

|  |  |  |  |  |  |  |  XR=1: 1

|  |  |  |  |  |  XQ=1:

|  |  |  |  |  |  |  XM=0: 1

|  |  |  |  |  |  |  XM=1: 0

|  |  |  |  |  XH=1: 0

Pre-Pruned Accuracy

--------------------------------

Number of training instances =  600

Number of training attributes =  21

Total Number of nodes in the tree =  265

Number of leaf nodes in the tree =  133

Accuracy of the model on the training dataset :  100.0 %

Number of validation instances =  2000

Number of validation attributes =  21

Accuracy of the model on the validation dataset before pruning :  74.3 %

Number of test instances =  2000

Number of test attributes =  21

Accuracy of the model on the testing dataset :  75.14999999999999 %

Pruning attempt  1 .....

Pruning attempt  2 .....

Pruning attempt  3 .....

Pruned Accuracy

--------------------------------

Number of training instances =  600

Number of training attributes =  21

Total Number of nodes in the tree =  219

Number of leaf nodes in the tree =  110

Accuracy of the model on the training dataset :  96.5 %

Number of validation instances =  2000

Number of validation attributes =  21

Accuracy of the model on the validation dataset after pruning :  74.9 %

Number of test instances =  2000

Number of test attributes =  21

Accuracy of the model on the testing dataset :  75.55 %