CS 6375.002 - Machine Learning - Spring '17

Assignment 1 - Feb 19, 2017

Problem:

To implement Decision Tree using ID3 algorithm with the following heuristics:

- ➢ Information Gain using Entropy
- ➢ Gain using Variance Impurity

Implementation Files:

File Name	Usage
Main.java	Main Driver Program to run
ReadFiles.java	To read values from Data Set.
Gain.java	Calculates IG using Entropy and VI
Tree.java	Tree structure and its methods
DecisionTree.java	To build the Decision Tree
Pruning.java	Implementation of Pruning Algorithm

Data Set:

Given Data Set Files	Purpose
Training Set	To construct Decision Tree
Test Set	To find accuracy after building the Tree
Validation Set	Tree is pruned using Validation Set to improve accuracy.

Results:

(The steps to Compile and Run are given in the Readme.txt file)

Before Pruning:

Data Set	Accuracy - Before Pruning	
	from IG using Entropy	**from Gain using Variance**
1	75.94999999999999	76.5
2	72.83333333333334	72.33333333333334

After Pruning:

Data Set	L value	K value	Accuracy - After Pruning	
			from IG using Entropy	**from Gain using Variance Impurity**
1	49	19	75.94	76.50
	14	2	76.35	76.5
	20	61	75.94	76.50
	127	4	76.44	76.55
	91	9	75.05	76.50
2	5	17	72.83	72.3
	125	12	73.0	71.0
	57	9	72.833	71.667
	201	4	72.5	72.166
	501	19	73.1667	71.8334

Conclusion:

The decision tree is implemented using the heuristics: variance impurity and entropy and the accuracy has been improved by pruning the tree.