Homework 1

September 12, 2018

1 Homework 1B

Build a data set, and use it to prove that decision tree built by greedy algorithm may not be the best desicion tree.

Theorem 1 *Decision tree built by greedy algorithm may not be the best desicion tree.*

Proof 1 *If there is no limitation for building the tree, all ways for generating the tree will finally lead to tree with the same information entropy. Therefore, we must introduce limitation to get different entropy.*

We bulid the data set as following, where A, B are attributes and X is class.

ID	A	B	X
1	0	0	0
2	0	0	0
3	0	1	0
4	0	1	1
5	1	0	1
6	1	0	1
7	1	1	1
8	1	1	0

When using greedy algorithm, the first division happened at attribute A, with a entropy reduction of log8-((1/4)log(4)+(3/4)log(4/3)).

While using attribute B to make the 1st node spilt, the entropy loss is log8-2((1/2)log(2)), which is less than the former one.

Then we can choose a number between the two losses, and make sure the tree will not grow when the entropy loss is less than the number.

In that case, the greedy algorithm will stop here. Yet non-greedy algorithm will not stop and spilt the new nodes, which will generate the less entropy in that dataset.

Therefore the data set above can prove the theorem.