

HOMEWORK EXAMPLE FOR DATA MINING (cse634)

CSE352

HOMEWORK 4 (Decision Tree Learning)

TRAINING DATA SET FOR THE HOMEWORK: Class Attribute: **Buys Computer**

Age	Income	Student	Credit Rating	Buys Computer
<=30	high	No	Fair	No
<=30	high	No	Excellent	No
31...40	high	No	Fair	Yes
>40	medium	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	low	Yes	Excellent	No
31...40	low	Yes	Excellent	Yes
<=30	medium	No	Fair	No
<=30	low	Yes	Fair	Yes
>40	medium	Yes	Fair	Yes
<=30	medium	Yes	Excellent	Yes
31...40	medium	No	Excellent	Yes
31...40	high	Yes	Fair	Yes
>40	medium	No	Excellent	No

Problem 1

Use the Training Data to create two decision trees:

1. Tree 1 - with majority voting (if needed);
2. Tree 2 - with general majority voting , i.e. majority voting at any node of your choice.
3. For the tree 1. use INCOME as root attribute, and nodes attributes of your choice;
4. For the tree 2. use CREDIT RATING as the root attribute, and nodes attributes of your choice;

TEST DATA SET

Obj	Age	Income	Student	Credit_Rating	Class
1	<=30	High	Yes	Fair	Yes
2	31...40	Low	No	Fair	Yes
3	31...40	High	Yes	Excellent	No
4	>40	Low	Yes	Fair	Yes
5	>40	Low	Yes	Excellent	No

6	≤ 30	Low	No	Fair	No
---	-----------	-----	----	------	----

Problem 2

Create test data sets for your sets rules corresponding to trees 1. and 2. that guarantees 100% predictive accuracy.

Problem 3

Compute the predictive accuracy of the set of rules in the lecture notes (with respect of the TEST Dataset from Problem 1.

Problem 4

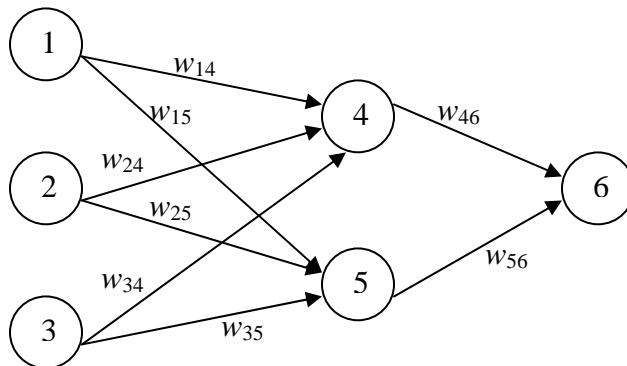
1. Create a classification data base of 20 records with 6 attributes (non- class) and 3 classes. At least 3 attributes must have continuous numerical values.
 2. Create 2 sets of data D1 and D2 using 2 discretization methods.
 3. Create 2 classifiers based on D1 and D2, respectively.
- Describe each step in the process.
Write justification of all methods you decided to use in the process.

Problem 4 Learning with Neural Networks

Given two records (Training Sample)

a_1	a_2	a_3	Class
0.5	0	0.2	1
0	0.3	0	1

1. Use the Network below to evaluate a passage of TWO EPOCHS.



Learning Rate $\ell = 0.7$

REMEMBER: YOU HAVE TO SET YOUR INITIAL WEIGHTS AND BIASES
RANDOMLY; DON'T USE THE SET-UP FROM THE
EXAMPLE.

2. Write you're the terminating conditions for your network
 3. Write a condition for success; i.e. how you decide that the record is well classified.
- b
-