

Assignment 2

Artificial Intelligence 2009 Track 1

You are working as expert at the analysis department in a large bank. The bank is manually evaluating all applications for loans, but would now like to automate the process. One idea that has come up is Decision Trees. Your task is to examine how this technique can be used to automate the evaluation of loan applications. The result of the examination should be in the form of a report that can be used as decision support for the board of the bank.

Guidelines for the report:

- Write a short introduction about the bank application.
- Write a brief description of how DTs work.
- Show examples of how DTs works on real data (see next page and data file):
 - Split the data into training data and test data.
 - Label the nodes in the decision trees with real names, not A15, A7 etc. in the report.
 - Run Cross Validation and check how much the generated trees differ from each other.
- Examine how the generation of the trees can be tuned to reflect the fact that we prefer to be overly hard in the judgment of people who apply for credit. Show with an example.
- How can the usage of the program be automated? I.e. can we avoid the Windows interface?
- How can one decide if DT is better than the present manual system? Suggest a procedure.

The decision tree program See5 is available on the PCs in the lab (Startmenu -> Programming -> See5). In addition there is a free stripped demo version and a tutorial available at <http://www.rulequest.com/>

Data for AI assignment 2

Dataformat:

A1: b,a.
A2: continuous.
A3: continuous.
A4: u, y, l, t.
A5: g, p, gg.
A6: c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff.
A7: v, h, bb, j, n, z, dd, ff, o.
A8: continuous.
A9: t,f.
A10: t,f.
A11: continuous.
A12: t,f.
A13: g, p, s.
A14: continuous.
A15: continuous.

A16: + or - (classification)

3 examples in human readable format:

sex	M	M	F
age	24	?	58
mean time at addresses	2.75	0.625	21
home status	owner	owner	owner
home telephone	given	given	given
current occupation	creative etc	laborer	guard etc
current job status	private sector	private sector	self-employed
mean time with employers	2.25	0.25	10
other investments	t	f	t
bank account	t	f	t
time with bank	6	-	13
liability reference	f	f	f
account reference	given	given	given
monthly housing expense	?	380	0
savings account balance	600	2010	6700

Sample Data (more data in file Assignment 2.dat)

b,22.08,0.83,u,g,c,h,2.165,f,f,0,t,g,00128,0,+
b,29.92,1.835,u,g,c,h,4.335,t,f,0,f,g,00260,200,+
a,38.25,6,u,g,k,v,1,t,f,0,t,g,00000,0,+
b,48.08,6.04,u,g,k,v,0.04,f,f,0,f,g,00000,2690,+
a,45.83,10.5,u,g,q,v,5,t,t,07,t,g,00000,0,+
b,36.67,4.415,y,p,k,v,0.25,t,t,10,t,g,00320,0,+
b,28.25,0.875,u,g,m,v,0.96,t,t,03,t,g,00396,0,+
a,23.25,5.875,u,g,q,v,3.17,t,t,10,f,g,00120,245,+
b,21.83,0.25,u,g,d,h,0.665,t,f,0,t,g,00000,0,+
a,19.17,8.585,u,g,cc,h,0.75,t,t,07,f,g,00096,0,+
b,25.00,11.25,u,g,c,v,2.5,t,t,17,f,g,00200,1208,+
b,23.25,1,u,g,c,v,0.835,t,f,0,f,s,00300,0,+
a,47.75,8,u,g,c,v,7.875,t,t,06,t,g,00000,1260,+
a,27.42,14.5,u,g,x,h,3.085,t,t,01,f,g,00120,11,+
a,41.17,6.5,u,g,q,v,0.5,t,t,03,t,g,00145,0,+