rule based classifier

rule= (Status= single) → No

Tid	Refund	Marital Status	Taxable Income	Class
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

coverage= 40%

accuracy= 50%

mutually exclusive rule= every record is covered at most one rule

↳ when rules are not → a record may trigger more than one rule ↱ order rules decide according to ranking

exhaustive rule= each record is covered by at least one rule

↳ when rules are not → a record may not trigger any rules → use default

building classification rules

direct method: sequential covering=

↳ start with empty rule → learn one rule → remove covered records by rule → repeat

R0: {} => class (initial rule)
R1: {A} => class (rule after adding conjunct)

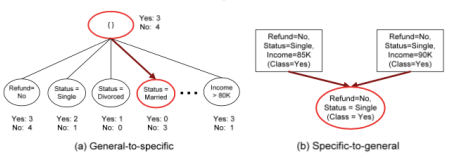
$$Gain(R_0, R_1) = p_1 \times [\, log_2\left(\frac{p_1}{p_1 + n_1}\right) - log_2\left(\frac{p_0}{p_0 + n_0}\right)]$$

p_0: number of positive instances covered by R0
n_0: number of negative instances covered by R0
p_1: number of positive instances covered by R1
n_1: number of negative instances covered by R1

Measure for pruning: v = (p-n)/(p+n)

- p: number of positive examples covered by the rule in the validation set
- n: number of negative examples covered by the rule in the validation set

↳ while there is a increase in gain, continue to add rules



indirect methods = decision trees, C4.5 rules (trying to prune decision tree for rules)