

Classification Rules

- Classify records by using a collection of “if...then...” rules
- An IF-THEN rule is an expression of the form IF “condition”, THEN “Conclusion”.
- Eg:R1:if age=“youth” and student=“yes” Then Buys_ computer=“yes”
- The “IF” part(left side) of a rule is known as the rule antecedent/precondition.
- The “THEN” part(right side) is the consequent.

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
human	warm	yes	no	no	mammals
python	cold	no	no	no	reptiles
salmon	cold	no	no	yes	fishes
whale	warm	yes	no	yes	mammals
frog	cold	no	no	sometimes	amphibians
komodo	cold	no	no	no	reptiles
bat	warm	yes	yes	no	mammals
pigeon	warm	no	yes	no	birds
cat	warm	yes	no	no	mammals
leopard shark	cold	yes	no	yes	fishes
turtle	cold	no	no	sometimes	reptiles
penguin	warm	no	no	sometimes	birds
porcupine	warm	yes	no	no	mammals
eel	cold	no	no	yes	fishes
salamander	cold	no	no	sometimes	amphibians
gila monster	cold	no	no	no	reptiles
platypus	warm	no	no	no	mammals
owl	warm	no	yes	no	birds
dolphin	warm	yes	no	yes	mammals
eagle	warm	no	yes	no	birds

- R1: (Give Birth = no) $\dot{\cup}$ (Can Fly = yes) [®] Birds
- R2: (Give Birth = no) $\dot{\cup}$ (Live in Water = yes) [®] Fishes
- R3: (Give Birth = yes) $\dot{\cup}$ (Blood Type = warm) [®] Mammals R4: (Give Birth = no) $\dot{\cup}$ (Can Fly = no) [®] Reptiles
- R5: (Live in Water = sometimes) [®] Amphibians
-

- In the rule antecedent, the condition consist of one or more attribute test(eg:-age="youth" and student="yes") that are logically AND ed.
- The rule consequent contains a class prediction(In this case, we are predicting whether a customer will buy a computer)
- R1 can also be written as:

R1:(age="youth")^(student=yes)=>(buy_s computer=yes)

- **Difference between rules and tree!!!!**

- ✓ The tree has a implied order in which the splitting is performed. Rule have no order.
- ✓ A tree is created based on looking at all classes. When generating rules, only one class must be examined at a time.

Rule Coverage and Accuracy

- **Coverage of a rule**: Fraction of records that satisfy the antecedent of a rule.

$$\text{coverage}(R) = n_{\text{covers}} / |D| \quad /* \text{ D: training data set } */$$

- **Accuracy of a rule** : Fraction of records that satisfy both antecedent and consequent of rule.

$$\text{accuracy}(R) = n_{\text{correct}} / n_{\text{covers}}$$

- If a rule is satisfied by x, the rule is said to be triggered.

<i>Tid</i>	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

(Status=Single) → No

Coverage = 40%, Accuracy = 50%

Problems

- What if more than one rule is satisfied and they belong to different classes?
- What if no rule is satisfied by X.

□ **Conflict Resolution Strategy:-**

- **Size ordering**
- **Rule Ordering**

- Size ordering:-

- ✓ It gives highest priority to the triggering rule that have the “toughest” requirement.

- ✓ Toughness is measured by the rule antecedent size

- ✓ Eg: $A1 \wedge A2 \rightarrow R1$

$(A1 \wedge A2) \wedge (A3 \wedge A4) \rightarrow R2$

- Rule Ordering:

- ✓ Ordering may be class-based or rule based.

- ✓ Class based:

- **Class based:**

- ❖ The class are sorted in order of decreasing “importance”

- ❖ All the rules for the most frequent class comes, then the next frequent and so on.

- ❖ Alternatively, they may be sorted based on misclassification cost per class.

