

Data Mining, CSCI 347, Fall 2019

Covering Rules, Sept. 25

Rules:

- Can be seen as “nuggets” of knowledge
- IF antecedent THEN consequence
- Antecedent and consequence can be attribute/value pairs, or several attribute/value pairs ANDed together (usually not ORed)

Two types of rules

- Classification rules – Supervised learning, use rules to predict the class value
- Association rules – Unsupervised learning, use rules to show relationships

Rather than looking at what attribute to split on, start with a particular class. Class by class, develop rules that “cover” the class.

Coverage (support) of a rule – number of instances predicted correctly by the rule, p

Accuracy (confidence) of a rule – coverage divided by the number of instance to which the rule applies, p/t

Algorithm for constructing classification rules:

```
For each class C
  Initialize E to the instance set
  While E contains instances in class C
    Create a rule R with an empty left-hand side that predicts class C
    Until R is perfect (or there are no more attributes to use) do
      For each attribute A not mentioned in R, and each value v,
        Consider adding the condition A = v to the left-hand side of R
        Select A and v to maximize the accuracy p/t
        (break ties by choosing the condition with the largest p)
      Add A = v to R
    Remove the instances covered by R from E
```