Rule-based Classification

Example ...

The Vertebrate dataset

<table>
<thead>
<tr>
<th>Name</th>
<th>Black Type</th>
<th>Can Fly</th>
<th>Live in Water</th>
<th>Give Birth</th>
<th>Blood Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>eel</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>python</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>dolphin</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>eagle</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>bat</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>komodo</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>frog</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>python</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>whale</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Terminology

Rule set: \( R = (r_1 \lor r_2 \lor \ldots \lor r_n) \)

Rule: \( r_i : (\text{Condition}) \rightarrow \text{C}_i \)

- Condition: \((A_1 \text{ op } v_1) \land (A_2 \text{ op } v_2) \land \ldots \land (A_k \text{ op } v_k)\)
- Rule antecedent, predition
- \( \text{Conjunct} : (A \text{ op } v), \text{op} \in \{\oplus,\ominus,\odot,\leq,\geq\}\)
- \( \text{C}_i \)
  - Class label
  - Rule consequent

Coverage and Accuracy

A rule \( r \) covers a record \( x \) if the precondition of \( r \) matches the attributes of \( x \)

- A.K.A. \( r \) is fired/triggered by \( x \)
- \( \text{Coverage}(r) = |A| / |D| \)
- \( |A| \): # of records covered by \( r \)
- \( \text{Accuracy}(r) = |A \cap y| / |A| \)
- \( |A \cap y| \): # of records that satisfy both the antecedent and consequent of \( r \)
- Example
  - coverage and accuracy of \( r_3 \)?

How a Rule-based Classifier Works

Lemur: ??
Turtle: ??
Dogfish shark: ??
Two Properties of a Rule-based Classifier

- Exhaustive Rules
  - Every combination of the attribute values is covered by at least one rule
- Mutually Exclusive Rules
  - No two rules are triggered by the same record

Make a Rule Set

- Exhaustive/Mutually Exclusive
  - Default rule: \((\) \(\rightarrow c_d)\)
  - Ordered rules
    - Quality-based ordering
    - Class-based ordering
  - Unordered rules
    - Majority votes
      - Weighted by the rule's accuracy

Sequential Covering Algorithms

- Order the classes \(\{c_1, c_2, \ldots, c_k\}\)
- For each class \(c_i, i \leq k\)
  - Find the best rule \(r\) for \(c_i\)
  - Remove the records covered by \(r\)
  - Add \(r\) to the rule list
  - Repeat until some stop condition is met
- Add a default rule \((\) \(\rightarrow c_k)\)

Sequential Covering Example

Ordering Classes and Rules

- Class ordering
  - Based on frequency
- Rule ordering
  - Based on classes
  - Based on quality of the rules

Rule Growing

- From general to specific
  - Start with \((\) \(\rightarrow c_i)\)
  - Greedily add one conjunct at a time
- From specific to general
  - Start with any positive record
  - Greedily remove one conjunct at a time
  - Augmented by beam search with k best candidates
Rule Growing Example (a)

Rule Growing Example (b)

Rule Evaluation

Decide which conjunct should be added (or removed)

Rule Evaluation Example

A training set contains 60 records in class $c_1$ and 100 records in class $c_2$
Comparing two rules:
- $r_1$: covers 50 $c_1$ and 5 $c_2$
- $r_2$: covers 2 $c_1$ and 0 $c_2$

Rule Evaluation Measure (a)

Likelihood Ratio:
$$ R(r) = 2 \sum_{i=1}^{k} f_i \log \left( \frac{f_i}{e_i} \right) $$

$e_i$: observed # of class $i$ records covered by $r$
$e_i$: expected # of class $i$ records covered by $r$

Rule Evaluation Measure (b)

FOIL's information gain:
$$ FGain(r) = n'_i \times \left( \log_2 \frac{n'_i}{n_i} - \log_2 \frac{n_i'}{n} \right) $$

<table>
<thead>
<tr>
<th># of records covered by $x$</th>
<th># of correct records covered by $x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before rule growth</td>
<td>$n$</td>
</tr>
<tr>
<td>After rule growth</td>
<td>$n'_i$</td>
</tr>
</tbody>
</table>
Stop Conditions

- Stop growing a rule
- Stop adding a rule for class $c_i$
  - Minimum Description Length (MDL)

Rule Pruning

- Similar to post-pruning of decision trees
- Remove a conjunct if the accuracy rate improves based on a validation set

Indirect Rule Extraction

- Using decision tree
  - Rule generation
  - Exhaustive?? Mutually Exclusive??
- Using association rule mining
  - Find association rules in the form of $A \rightarrow c_i$
  - Select a subset of the rules to form a classifier
    - Sort the rules based on confidence, support, and length
    - Add to a rule list one at a time
    - Add a default rule

Readings

- Textbook Chapter 8.4