EXERCISES

10.1. Consider a sequential covering algorithm such as CN2 and a simultaneous covering algorithm such as ID3. Both algorithms are to be used to learn a target concept defined over instances represented by conjunctions of ***n*** boolean attributes. If ID3 learns a balanced decision tree of depth d, it will contain $2^d - 1$ distinct decision nodes, and therefore will have made $2^d - 1$ distinct choices while constructing its output hypothesis. How many rules will be formed if this tree is re-expressed as t a disjunctive set of rules? How many preconditions will each rule possess? How many distinct choices would a sequential covering algorithm have to make to learn this same set of rules? Which system do you suspect would be more prone to overfitting if both were given the same training data?

Ans.

(a) The number of leaf nodes of a balanced decision tree is 2^{d-1}, so the number of rules is 2^{d-1}.

(b) d.

(c) $d*2^{d-1}$.

(d) Both algorithms have the problem of overftting because they will find the most specific hypothesis fit for the noisy training data. In my opinion, decision tree is more prone to overfitting because the expanding of a tree node has the influence on all descendant nodes which leads to overfitting, where in sequential covering algorithm, the influence of overfitting is limited in single hypothesis.

10.2. Refine the LEARN-ONE-RULE Algorithm of Table 10.2 so that it can learn rules whose preconditions include thresholds on real-valued attributes (e.g., temperature > 42). Specify your new algorithm as a set of editing changes to the algorithm of Table 10.2. Hint: Consider how this is accomplished for decision tree learning.

Ans.

1. Generate the next more specific Candidate_hypotheses

Calculate the best threshold for every continuous-valued attribute. (similar to decision tree)

All_constraints = the set of all constraints of such form:

the form (a=v) where a is a member of a discrete-valued attribute, and v is a value of a that occurs in the current set of Examples

the form (a>v) or (a $\leq$v) where a is a member of a continuous-valued attribute, and v is the threshold of a.

10.5. Apply inverse resolution in propositional form to the clauses $C = A \vee B$, $C1 = A \vee B \vee G$. Give at least two possible results for C2.

Ans. $C2 = \neg G$

$C2 = A \vee \neg G$

$C2 = B \vee \neg G$

$C2 = A \vee B \vee \neg G$

10.6. Apply inverse resolution to the clauses $C = R(B, x) \vee P(x, A)$ and $C1 = S(B, y) \vee R(z, x)$. Give at least four possible results for C2. Here A and B are constants, x and y are variables.

Ans. $C2 = P(x, A) \vee \neg S(B, y)$. $\theta = \{z/B\}$

$C2 = P(x, A) \vee \neg S(z, y)$. $\theta = \{z/B\}$

$C2 = P(x, w) \vee \neg S(z, y)$. $\theta = \{z/B, w/A\}$

$C2 = P(x, w) \vee \neg S(B, y)$. $\theta = \{z/B, w/A\}$

$C2 = P(x, A) \vee \neg S(B, y) \vee R(B, x)$. $\theta = \{z/B\}$

$C2 = P(x, A) \vee \neg S(z, y) \vee R(z, x)$. $\theta = \{z/B\}$