MACHINE LEARNING

1. Concept Learning

Target concept: girls who like Simon
Target function:
data: girls described by: hair, body, likes Simon, pose, smile, smart

Idea is to find best hypothesis to minimize *error rate*
Finite search space since data items are finite
general-to-specific ordering: - h1 *precedes* h2 off for all data, h1(data) gives positive classification ¡- h2(data) gives positive classification so

$$h1 >_g h2$$

- h2 and h2 have equal generality if exists data item if h1(d) = I -¿ h2(d)= I or h2(d) = I -¿ h1(d) = I, then

$$h1 =_g h2$$

If two hypotheses have the same number of defined attributes, they are of the same generality.

Find-S Algorithm

- Initialize hypothesis h to most specific hypothesis
- iterate over all data $d \in D$: iterate over all attributes in h: if the attribute is not satisfied by d replace the attribute s.t. $h' >_g h, h \leftarrow h'$

Lots of limitations of this algorithms as could output only one of many equally valid hypotheses.

Candidate Elimination Algorithms

- general-to-specific ordering of hypotheses
- two-sided approach to converging to solution
- specific boundary (initialize all attributes) works with positive examples and eliminates those that don't fit
- general boundary (let all attributes vary) works with negative examples and eliminates those that don't' fit

2. Decision Trees

Used for discrete approximation: helpful when *if-then* classification is needed. Task scheduling is a typical problem solved by decision trees. To lean using decision trees, use greedy search through space of possible solutions. Algorithm:

1: perform statistical test to determine how well the attribute classifies training data
2: best attribute forms root of tree

3: descendant node in each branch will be determined by attributes of the node; split data according to those attributes and continue

ID3 algorithm uses *information gain*:

$$E(S) \equiv -\sum_{v} p_v \ln(p_v)$$

where $v \in \{1 \ldots n\}$ and then

$$IG(S, A) = E(S) - \sum_{v \in values(A)} \left(\frac{|S|}{|S_v|} \right) E(S_v)$$

Iteratively, we take the value of the attribute that has the best explanatory power (information gain) and make that part of the decision tree deterministic. The other values will be classified again with subsets of the data set, and the process repeats itself.