CONCEPT LEARNING

How would you define: "A good day to go beach"?

=> Based on their everyday experiences humans easily learn abstract concepts - how computers can represent (model) and learn them?

Assumption: Concepts are learned from pos and neg examples.

Example: <slide>

Representation => model: <slide> {how to program?}

Example: "Joni enjoys his favourite sport on cold days with high humidity"
Give using the EnjoySport attributes

<?, Cold, High, ?, ?, ?>

=> we have a model!

1. Learning by search

Brute force: test all possible combinations

Example How many for EnjoySport?

possible combinations: $2 \cdot 2 \cdot 2 \cdot 1 \cdot 2 \cdot 2 = 32$

Hypotheses (+?, $\emptyset$): $4 \cdot 4 \cdot 4 \cdot 3 \cdot 4 \cdot 4 = 3072$

EnjoySport still doable, but easily explodes!

=> Faster method needed!

1.1 General & specific hypothesis

Example: Which is more general?

$h_1 = \langle Sunny, ?, ?, Strong, ?, ? \rangle$

$h_2 = \langle Sunny, ?, ?, ?, ?, ? \rangle$

h_2 covers h_1

Formal definition

$h_j \geq_g h_k$ iff

$(\forall x \in X)\ [(h_k(x) = 1) \rightarrow (h_j(x) = 1)]$ {how to program?}

2. FIND-S <slide>

Example Run FIND-S for EnjoySport

1° $h \leftarrow \langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$

2° $h \leftarrow \langle Sunny, Warm, Normal, Strong, Warm, Same \rangle$

3° $h \leftarrow \langle Sunny, Warm, ?, Strong, Warm, Same \rangle$

4° $h \leftarrow \langle Sunny, Warm, ?, Strong, ?, ? \rangle$

guaranteed to find the most specific consistent with the positive examples (also req if no errors)

Problems: <slide>

3. Version spaces <slide> {how to program Consistent()}

LIST-THEN-ELIMINATE: <slide>

What is wrong? => enormous memory needed!

3.1 Efficient VS representation

Example Find all consistent hypotheses of EnjoySport

FIND-S: S: $\langle Sunny, Warm, ?, Strong, ?, ? \rangle$ 1

$\langle ?, Warm, ?, Strong, ?, ? \rangle$ 2 $\quad \langle Sunny, ?, ?, Strong, ?, ? \rangle$ 3 $\quad \langle Sunny, Warm, ?, ?, ?, ? \rangle$ 4

G: $\langle ?, Warm, ?, ?, ?, ? \rangle$ 5 $\quad \langle Sunny, ?, ?, ?, ?, ? \rangle$ 6

We need to store only S and G!! <slide>

4. CANDIDATE-ELIMINATION algorithm <slide>

Example Walkthrough of EnjoySport <overheads>

Example Classify the following examples:

							h_1	h_2	h_3	h_4	h_5	h_6
A	Sunny	Warm	Normal	Strong	Cool	Change	1	1	1	1	1	1
B	Rainy	Cold	Normal	Light	Warm	Same	0	0	0	0	0	0
*C	Sunny	Warm	Normal	Light	Warm	Same	0	0	0	1	1	1
**D	Sunny	Cold	Normal	Strong	Warm	Same	0	0	1	0	0	1

* Random / majority of EnjoySport

** More votes on 0

Which would be the most beneficial to know?