CS1306 - MACHINE LEARNING - WEEK 3

Ex.No.3 **Date: 19.8.24**

CANDIDATE-ELIMINATION and Find -S Learning Algorithm

Training Dataset :

['sunny', 'warm', 'normal', 'strong', 'warm', 'same', 'yes']

['sunny', 'warm', 'high', 'strong', 'warm', 'same', 'yes']

['rainy', 'cold', 'high', 'strong', 'warm', 'change', 'no']

['sunny', 'warm', 'high', 'strong', 'cool', 'change', 'yes']

CANDIDATE- ELIMINTION algorithm using version spaces Training

CANDIDATE-ELIMINATION Learning Algorithm

The CANDIDATE-ELIMINTION algorithm computes the version space containing all hypotheses from H that are consistent with an observed sequence of training examples.

1. Initialize G to the set of maximally general hypotheses in H
2. Initialize S to the set of maximally specific hypotheses in H
3. For each training example d, do

• If d is a positive example

- Remove from G any hypothesis inconsistent with d
- For each hypothesis s in S that is not consistent with d
  - Remove s from S
  - Add to S all minimal generalizations h of s such that
  - h is consistent with d, and some member of G is more general than h
- Remove from S any hypothesis that is more general than another hypothesis in S

• If d is a negative example

- Remove from S any hypothesis inconsistent with d
- For each hypothesis g in G that is not consistent with d

- Remove g from G
- Add to G all minimal specializations h of g such that

- h is consistent with d, and some member of S is more specific than h

- Remove from G any hypothesis that is less general than another hypothesis in G

Algorithmic steps:

Initially : G = [[?, ?, ?, ?, ?, ?], [?, ?, ?, ?, ?, ?], [?, ?, ?, ?, ?, ?],
[?, ?, ?, ?, ?, ?], [?, ?, ?, ?, ?, ?], [?, ?, ?, ?, ?, ?]]
S = [Null, Null, Null, Null, Null, Null]

For instance 1 : <'sunny','warm','normal','strong','warm ','same'> and positive output.
G1 = G
S1 = ['sunny','warm','normal','strong','warm ','same']

For instance 2 : <'sunny','warm','high','strong','warm ','same'> and positive output.
G2 = G
S2 = ['sunny','warm',?,'strong','warm ','same']

For instance 3 : <'rainy','cold','high','strong','warm ','change'> and negative output.
G3 = [['sunny', ?, ?, ?, ?, ?], [?, 'warm', ?, ?, ?, ?], [?, ?, ?, ?, ?, ?],
[?, ?, ?, ?, ?, ?], [?, ?, ?, ?, ?, ?], [?, ?, ?, ?, ?, 'same']]
S3 = S2

For instance 4 : <'sunny','warm','high','strong','cool','change'> and positive output.
G4 = G3
S4 = ['sunny','warm',?,'strong', ?, ?]

At last, by synchronizing the G4 and S4 algorithm produce the output.
G = [['sunny', ?, ?, ?, ?, ?], [?, 'warm', ?, ?, ?, ?]
S = ['sunny','warm',?,'strong', ?, ?]

Instruction:
The training examples should be saved in csv file format. Enjoysport.csv and save in the same folder where this python file is saved

FIND-S Algorithm

1. Initialize h to the most specific hypothesis in H
2. For each positive training instance x
For each attribute constraint ai in h
If the constraint ai is satisfied by x
Then do nothing
Else replace ai in h by the next more general constraint that is satisfied by x
3. Output hypothesis h

Output:

The Given Training Data Set
['sunny', 'warm', 'normal', 'strong', 'warm', 'same', 'yes']
['sunny', 'warm', 'high', 'strong', 'warm', 'same', 'yes']
['rainy', 'cold', 'high', 'strong', 'warm', 'change', 'no']
['sunny', 'warm', 'high', 'strong', 'cool', 'change', 'yes']

The initial value of hypothesis:
['0', '0', '0', '0', '0', '0']
Find S: Finding a Maximally Specific Hypothesis
For Training Example No:0 the hypothesis is
['sunny', 'warm', 'normal', 'strong', 'warm', 'same']
For Training Example No:1 the hypothesis is
['sunny', 'warm', '?', 'strong', 'warm', 'same']
For Training Example No:2 the hypothesis is
'sunny', 'warm', '?', 'strong', 'warm', 'same']
For Training Example No:3 the hypothesis is
'sunny', 'warm', '?', 'strong', '?', '?']
The Maximally Specific Hypothesis for a given Training Examples:
['sunny', 'warm', '?', 'strong', '?', '?']

uestion 1:

For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm and Find -S to output a description of the set of all hypotheses consistent with the training examples.

Origin	*Manufacturer*	*Color*	*Decade*	*Type*	*Example Type*
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive
Japan	Toyota	Green	1980	Economy	Positive
Japan	Honda	Red	1990	Economy	Negative

Question 2: Implement Candidate Elimination and Find S algorithm and find the difference between the hypotheses generated.

example	*citations*	*size*	*inLibrary*	*price*	*editions*	*buy*
1	some	small	no	affordable	many	no
2	many	big	no	expensive	one	yes
3	some	big	always	expensive	few	no
4	many	medium	no	expensive	many	yes
5	many	small	no	affordable	many	yes

.

Spot Exercise :

Consider a concept learning problem where the data D, which concerns ancient Egyptian vases discovered in archeological excavations, is expressed as tuples of five attributes:

damaged, color, material, kingdom, markings.

Examples are classified as either valuable (+) or not valuable (-), and D consists of the following:

Number	Example	Class
1	<no, brown, marble, middle, hieroglyphics>	+
2	<no, white, sandstone, old, none>	-
3	<no, white, marble, new, hieroglyphics>	+
4	<yes, grey, slate, middle, hieroglyphics>	-
5	<no, brown, granite, middle, hieroglyphics>	+

Assume that all possible values of each attribute are represented in D above.

(a) What is the size of the hypothesis space searched by the candidate elimination algorithm (CEA) using the data D given above?

(b) Suppose the CEA has seen examples 1 and 2 only so far. Show its current specific boundary S_2 and general boundary G_2 for the version space.

(c) Show S_3 and G_3 after the CEA also sees example 3.

(d) Show S_5 and G_5 after the CEA also sees the final two examples 4 and 5.