CS1306 - MACHINE LEARNING - WEEK 3

Question 1:

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

CANDIDATE- ELIMINTION algorithm using version spaces Training Examples:

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Question 2:

For a given set of training data examples stored in a .CSV file, implement and demonstrate the Find-S algorithm

Example	Size	Color	Shape	Class/Label
1	Big	Red	Circle	No
2	Small	Red	Triangle	No
3	Small	Red	Circle	Yes
4	Big	Blue	Circle	No
5	Small	Blue	Circle	Yes

Question 3:

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

Step 1:

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

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 .