Assignment 03

Name : Kiran K. Patil Sem : 06
ID : 2110709004 Course : ML - Lab.

Aim : Implement Candidate Elimination Algorithm on the titanic dataset.

Theory : The candidate elemination algorithm incrementally builds the version space given a hypothesis space H and a set E of examples. The examples are added one by one, each example possibly shrinks the version space by removing the hypothesis that are incosistent with example. The candidate elimination algorithm does this by updating the general and specific boundry for each new example.

- You can consider this as an extended form of Find-s algorithms.
- Consider both positive and negative examples
- Actually, positive examples are used here as Find-s algorithm
  (Basically they are generalizing from the specification)

- while the negative example is specified from generalize form.

Terms used:

- Consept learning: consept learning is basically learning task of the machine learing.
- General hypothesis: Not specifying features to learn the machine.
- G = { 'pi', '?', '?', '?' ... } Number of attr.
- specific hypothesis: specifying features to learn machine (specific feature)
- S = { 'pi', 'pi', 'pi', ... }. Number of Pi depends on number of attributes.
- version space: It is intermediate of general hypothesis and specific hypothesis it is only just written and hypothesis but a set of all possible hypothesis based on training data set.

Algorithm:

Step 1: Load Data set.

step 2: Initialize General hypothesis and specific Hypothesis.

step 3: For each training example

Step 4: If example is positive example
if attribute value == hypothesis_value
Do nothing.
else:
replace attribute value with '?'
(Basically generalizing it)

Step 5: If example is Negative example:
Make generalize hypothesis more specific.

Operations / preprocessing :-

↓ Removed unnessesary columns

→ df.drop(['cabin','Name','PassengerId'], axis=1, inplace=true)

4] classified age as categorywise.
infant : 0-3, child : 10, Teenager : 18, Young Adult : 25,
Adult : 40, Elderly : 80.

5] cassified fare category
↳ General, second class, first class.

6] classified siblings count
↳ Low, medium, High.

7] Set the target attributes as survived / - 0, 1.

dataset is looking as :

8]

survived	class	class	Parch	Embarked	Age	Fare	sibsp
0	male	3	0	Q	Adult	Gen	Low
1	Female	3	0	S	Elderly	Gen	Low
0	male	2	0	Q	Elderly	Gen	Low
0	male	3	0	S	Adult	Gen	Low
1	female	3	1	S	Young Adult	Gen	Low.

9] Then applied the candidate elimination algorithm.
- specific_algo : ['?', 0, '?', 'male', ?, ?, ?, General]
- General algo : [?, ?, ?, ?, ?, ?, ?, ?], [? ? ? male, ? ? ? ?.

Conclusion : Thus we have applied candidate elimination algorithm on filtered dataset and it is observed that on filtered dataset we get consistant hypothesis