ML Assignment 02

Name: Kiran K. Patil
ID : 211070904
sem : 06
course: ML Lab

Aim: Implement and demonstrate the find-s - algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the traing data from a .csv file.

Theory :

What is find-s Algorithm in Machine learing?
In order to understand find-s algorithm, you need to have a basic idea of the following as consepts as swell.

1. Consept learning :
2. General hypothesis
3. Specific hypothesis.

1. consept learning :-
lets try to understand consept learning with an example. Most of the humans learning is based on past instances or expriences. For example, we are able to identify any type of vehical based on a certain features like make, model etc. that are

defined over a large set of features. These special features differentiates the set of cars trucks etc from the larger set of vehicals These features that defines the set of cars, trucks etc are known as concepts. So the concept learning can be formulated as a problem of searching through predefined space of potential hypothesis for the hypothesis that best fits the training example.

In simple word concept learning is infering a boolean-valued function from training examples of its input and output.

2. General hypothesis:
Hypothesis in general is an explanation of somthing. The hypothesis basically states that general relationship between the major variables. For example, a general hypothesis for ordering food would be I want a burger
$G = \{ '?', '?', '?', '?' \}$

3. Specific hypothesis:
The specific hypothesis fill in all the important details about the variables given in the general hypothesis. The more specific details into the examp given above would be "I want a cheeseburger with lot of lettuce.
$S = \{ 'ø', 'ø', 'ø', 'ø' \}$

FIND-S Algorithm:

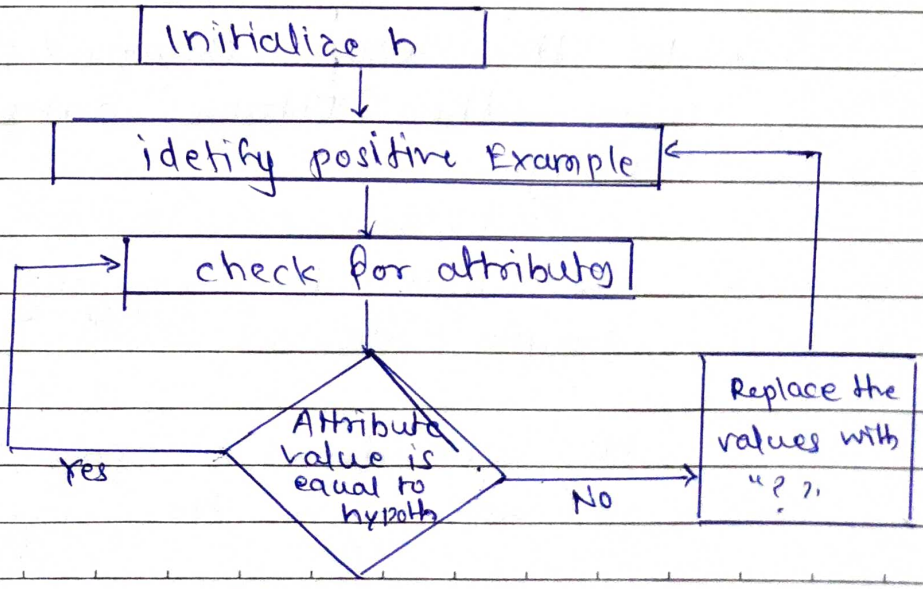
1. Initialize h to the most specific hypothesis in H
2. For each positive training instance x

   For each attribute constraint ai in h

   If the constraint 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.

How does it work?:



- Limitations of find-s Algorithm:

- There are few limitations of the find-s algorithm listed down below.

1. There is no way to determine if the hypothesis is consistent throughout the data.

2. Inconsistent training set can actully mislead the find-s algorithm, since it ignores the negative examples.

3. Find-s algorithm doesn't provide a backtracking technique to determine the best possible changes that could be done to improve the resulting hypothesis.

- Operations with the dataset.

- In this ~~exapte~~ example / Experiment we are using the "Titanic" dataset. available at kaggle.

- We will be applying find-s on the titanic dataset.

- If we directly apply the algorith then it will give the hypothesis <?, ?, ?, ?, ?, ?>

- So In order to get some meaningful hypothesis we have to filter the data accordingly.

1. Removing the unnessesary attributs / colums.

- for removing unnesary columns will use:
- data.drop ("passengerId', inplace = True, axis = 1)
- likewise we will remove, Passenge ID, Name, ticket, fare,
- Now we are having the dataset as follows

Survived	Pclass	Age	Pclass	sibsp	Parch	Cabin	Embarke
0	male	14	3	0	0	NaN	S
1	Female	30	3	0	0	NaN	Q
0	male	28	2	1	1	Nan	S
1	female	18.	3	0	0	Nan	C

2. The for above data-set when find-s algrith applied by reducing the num of rows.

- data = data [: 25]
- we get hypothesis as:

< 1, '?', 'Female', '?', '?', '0', '?', '?' >

Futthur tappleed some more functions.

3. Sort the data according to age.

```
data-frame.sort_values ('Age')
```

4. Drop the Null value records. (i.e.

```
valid_df = data_frame.dropna()
```

5. removing Rows / unnessary records.

```
sorted_df : valid_df [ :25]
```

6. Group by age (created the class intervals for the age)

Survived	places	sex	Age	Parch	Cabin	Emb.
1	1	female	30.0	1	C54	C
1	1	female	30.0	0	B46	S
0	1	male	30.0	0	B24	9
0	1	male	30.0	0	C31	S
0	1	male	40.0	0	B45	S

7. After applying Find-s on this filtered data got the hypothesis as follows.

< 1, 1, 'Female', 30, '?', '?', '?', '?' >

Conclusion: Thus from This Experiment we have applied Find-S Algorithm on the Titanic Dataset.

- without filtering when we had applied the algo. we got hypothesis as < ?, ?, ?, ?, ?, ?, ? >
- After applying some filters on the dataset and removing unnessesary columns we got the hypothesis. as
  < 1, '?', 'Female', '?', '?', '0', '?', ? >
- Then applied grouping and classified the data according to age group. and applied Find-s algorithm got the hypothesis as
  < 1, 1, 'Female', 30, '?', '?', '?', '?' >
- So from above observation in all three hypothesis we have got female as common attribut thus mostly females are survived.