**LAB EXERCISE – 7**

**Find – S Algorithm**

1. **Aim of the Experiment:**

Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file and generate the final specific hypothesis.

1. **Reference to Text book for Algorithms:**

Refer to Section 3.4.5 in Chapter 3 Basics of Learning Theory to understand the working of the algorithm.

**Listing 1:**

Sample Dataset Used: Table 3.2 in that chapter.

**Table 3.2:** Training Dataset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CGPA | Interactiveness | Practical Knowledge | Communication Skills | Logical Thinking | Interest | Job Offer |
| ≥9 | Yes | Excellent | Good | Fast | Yes | Yes |
| ≥9 | Yes | Good | Good | Fast | Yes | Yes |
| ≥8 | No | Good | Good | Fast | No | No |
| ≥9 | Yes | Good | Good | Slow | No | Yes |

**3. Python Program with Explanation:**

1. Import the package csv to work with .csv files.

import csv

2. The attribute values of each attribute in the given dataset are included in the list ‘attr\_values’.

attr\_values = [['g9','g8'],['Yes','No'],['Excellent','Good'],['Good','Bad'],

['Fast','Slow'],['Yes','No']]

3. The number of attributes are counted in the list with the len() function.

num\_attrs = len(attr\_values)

4. The most general hypothesis and the most specific hypothesis are printed.

print (" \n The most general hypothesis : ['?','?','?','?','?','?']\n")

print ("\n The most specific hypothesis : ['0','0','0','0','0','0']\n")

5. Create a dynamic array a[]. It is also called as a list.

a = []

6. Print the training dataset reading from the .csv file. First open the CSV file as a text file with open() function, which returns a file object. Pass the file object to the reader to read from the CSV file. Read each row and append to the list a[]. Print each row as appended to the list.

print("\n Training Dataset \n")

with open('Find-S csv.csv', 'r') as csvFile:

reader = csv.reader(csvFile)

for row in reader:

a.append (row)

print(row)

7. Print the initial value of the hypothesis with ‘0’ for the number of attributes in the dataset.

print("\n The initial value of hypothesis: ")

h = ['0'] \* num\_attrs

print(h)

8. Generate the hypothesis ‘h’ with the first training instance.

for j in range(0,num\_attrs):

h[j] = a[0][j];

9. Read subsequent training instances and generate hypothesis ‘h’.

print("\n Find S: Finding Maximally Specific Hypothesis\n")

for i in range(0,len(a)):

if a[i][num\_attrs]=='Yes':

for j in range(0,num\_attrs):

if a[i][j]!=h[j]:

h[j]='?'

else :

h[j]= a[i][j]

print(" Training Instance :{0} the hypothesis 'h' is ".format(i),h)

10. Print the final maximally specific hypothesis after reading all instances.

print("\n The Final Maximally Specific Hypothesis 'h' is :\n")

print(h)

**4. Complete Program:**

import csv

attr\_values = [['g9','g8'],['Yes','No'],['Excellent','Good'],['Good','Bad'],

['Fast','Slow'],['Yes','No']]

num\_attrs = len(attr\_values)

print (" \n The most general hypothesis : ['?','?','?','?','?','?']\n")

print ("\n The most specific hypothesis : ['0','0','0','0','0','0']\n")

a = []

print("\n Training Dataset \n")

with open('Find-S csv.csv', 'r') as csvFile:

reader = csv.reader(csvFile)

for row in reader:

a.append (row)

print(row)

print("\n The initial value of hypothesis: ")

h = ['0'] \* num\_attrs

print(h)

# The Hypothesis 'h' with the first training instance

for j in range(0,num\_attrs):

h[j] = a[0][j];

# The Hypothesis 'h' with subsequent training instances

print("\n Find S: Finding Maximally Specific Hypothesis\n")

for i in range(0,len(a)):

if a[i][num\_attrs]=='Yes':

for j in range(0,num\_attrs):

if a[i][j]!=h[j]:

h[j]='?'

else :

h[j]= a[i][j]

print(" Training Instance :{0} the hypothesis 'h' is ".format(i),h)

print("\n The Final Maximally Specific Hypothesis 'h' is :\n")

print(h)

**Output:**

Python 3.8.3 (tags/v3.8.3:6f8c832, May 13 2020, 22:37:02) [MSC v.1924 64 bit (AMD64)] on win32

Type "help", "copyright", "credits" or "license()" for more information.

>>>

=============== RESTART: C:/Users/ADMIN/pythonpgms/Find-S test.py ==============

The most general hypothesis : ['?','?','?','?','?','?']

The most specific hypothesis : ['0','0','0','0','0','0']

Training Dataset

['g9', 'Yes', 'Excellent', 'Good', 'Fast', 'Yes', 'Yes']

['g9', 'Yes', 'Good', 'Good', 'Fast', 'Yes', 'Yes']

['g8', 'No', 'Good', 'Good', 'Fast', 'No', 'No']

['g9', 'Yes', 'Good', 'Good', 'Slow', 'No', 'Yes']

The initial value of hypothesis:

['0', '0', '0', '0', '0', '0']

Find S: Finding Maximally Specific Hypothesis

Training Instance :0 the hypothesis 'h' is ['g9', 'Yes', 'Excellent', 'Good', 'Fast', 'Yes']

Training Instance :1 the hypothesis 'h' is ['g9', 'Yes', '?', 'Good', 'Fast', 'Yes']

Training Instance :2 the hypothesis 'h' is ['g9', 'Yes', '?', 'Good', 'Fast', 'Yes']

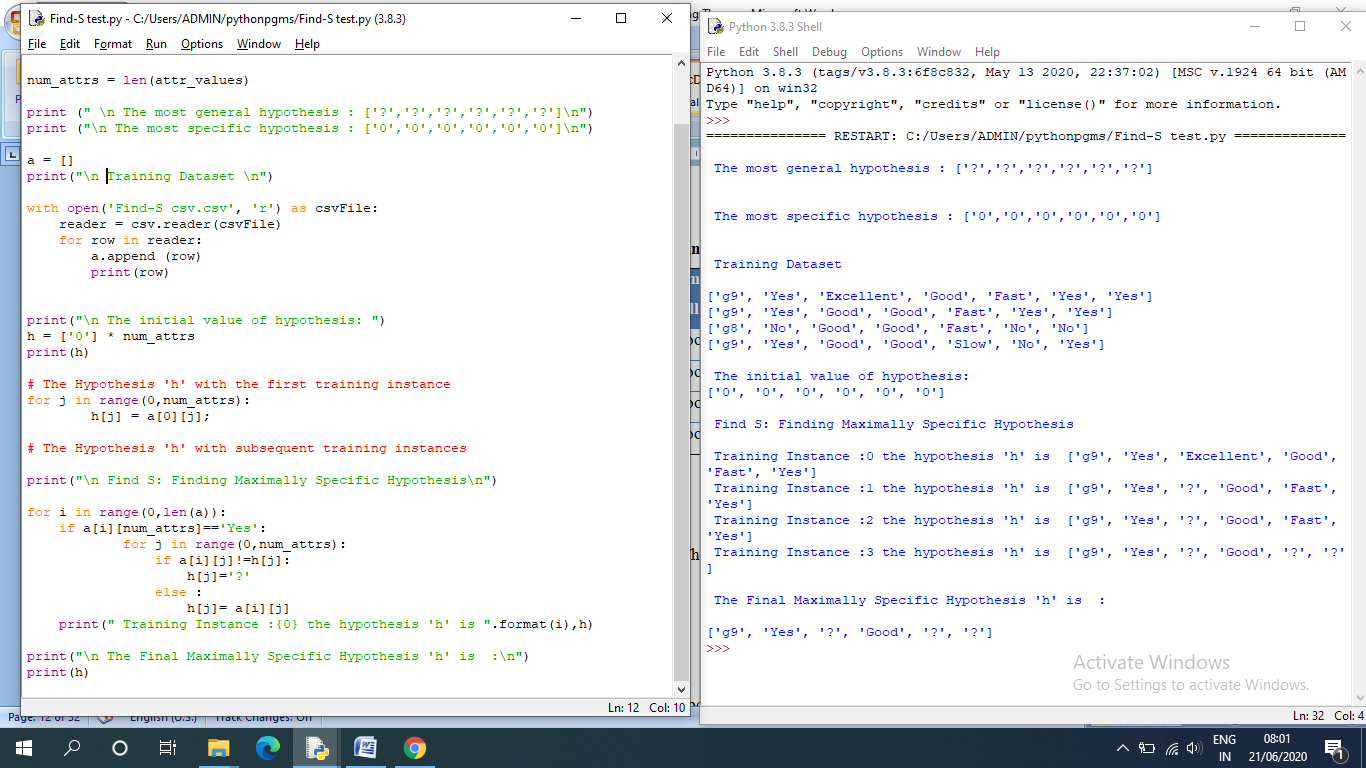
Training Instance :3 the hypothesis 'h' is ['g9', 'Yes', '?', 'Good', '?', '?']

The Final Maximally Specific Hypothesis 'h' is :

['g9', 'Yes', '?', 'Good', '?', '?']

>>>

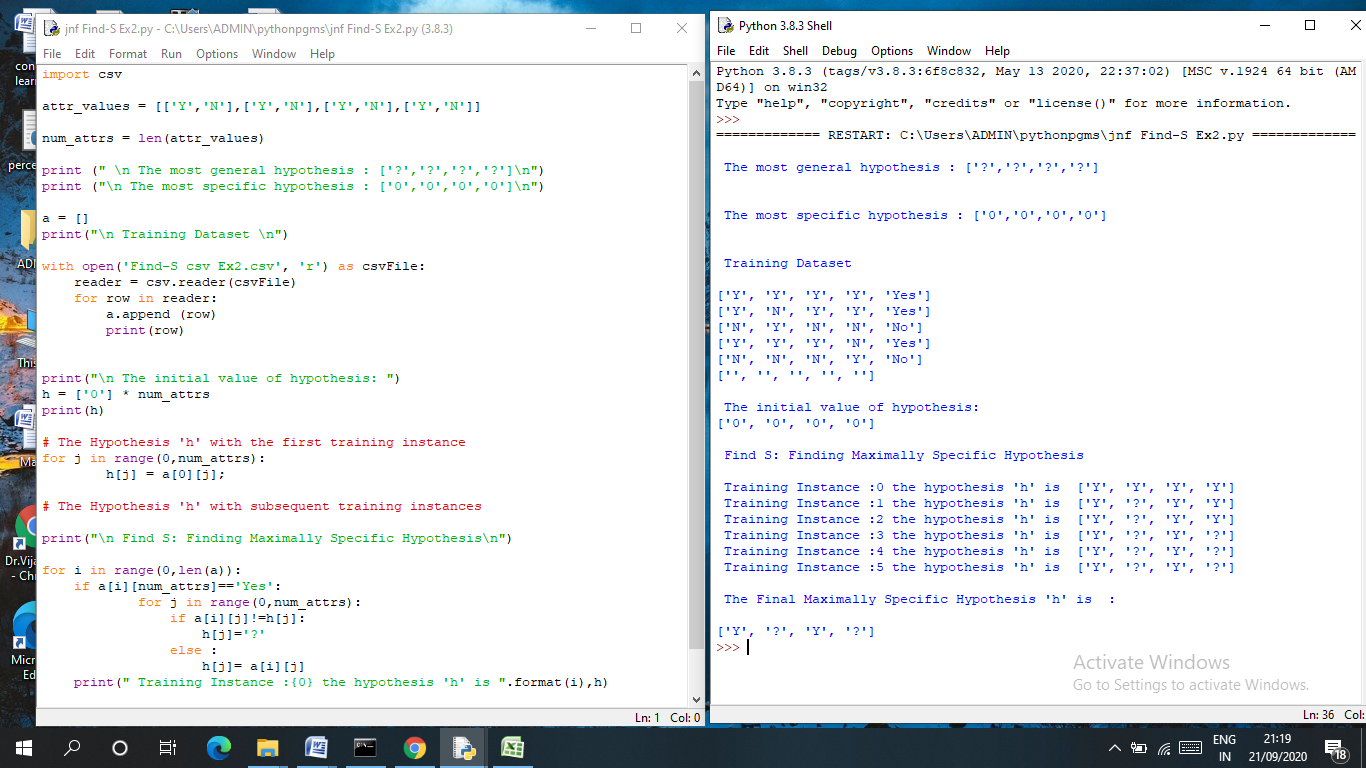
**Screenshot of the Output:**



**Listing 2:**

**Sample Training Instances**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| S.No. | Fever | Cough | Throat\_Pain | Body\_Pain | Covid19 |
| I1. | Y | Y | Y | Y | Positive |
| I2. | Y | N | Y | Y | Positive |
| I3. | N | Y | N | N | Negative |
| I4. | Y | Y | Y | N | Positive |
| I5. | N | N | N | Y | Negative |



**Listing 3:**

**Table 3.6: Sample Training Instances**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| S.No. | Fever | Cough | Throat\_Pain | Body\_Pain | Covid19 |
| I1. | N | Y | Y | Y | Positive |
| I2. | Y | Y | Y | Y | Positive |
| I3. | Y | N | Y | Y | Positive |
| I4. | N | Y | N | N | Negative |
| I5. | Y | Y | Y | N | Positive |
| I6. | N | N | N | Y | Negative |
| I7. | N | N | N | N | Negative |

