Sardar Patel Institute of Technology, Mumbai
Department of Electronics and Telecommunication Engineering
B.E. Sem-VII (2019-2020)
ETEL71A - Machine Learning and AI

Experiment: Find-S Algorithm

Name: Abhinav Pallayil **Date:** 13/09/2022
UID: 2019120045

Objective/Aim: To explore how Find-S algorithm is used for finding the most specific hypothesis based on a given set of training data samples.

Outcomes:

1. Representation of hypothesis
2. Apply Find-S algorithm on the given data to get the most specific hypothesis
3. Interpret the output of Find-S

System Requirements: Linux OS with Python and libraries or R or windows with MATLAB

Problem Statement: Using the Find-S Algorithm, determine whether a person is smart or not.

Data Set Link:

Sr.No	hair	body	glasses	pose	smile	smart
1	blond	thin	yes	arrogant	toothy	no
2	brown	thin	no	natural	pleasant	yes
3	blond	plump	yes	goofy	pleasant	no
4	black	thin	no	arrogant	none	no
5	blond	plump	no	natural	toothy	yes
6	black	thin	no	natural	pleasant	yes
7	blond	thin	yes	goofty	none	no
8	blond	plump	no	natural	pleasant	yes
9	blond	thin	no	natural	toothy	no
10	black	thin	no	natural	pleasant	no
11	blond	thin	yes	arrogant	pleasant	no
12	black	thin	no	natural	none	yes
13	blond	plump	no	natural	none	yes
14	blond	plump	yes	arrogant	toothy	no
15	blond	thin	no	natural	toothy	yes

Dataset Description:
Number of Instances: 15

Number of Attributes: 5

Attribute Information:

1. hair: 3 (black, blond, brown)
2. body: 2 (plump, thin)
3. glasses: 2 (yes, no)
4. pose: 3 (arrogant, natural, goofy)
5. smile: 3 (pleasant, toothy, none)

Target- smart: 2 (yes, no)

Algorithm:
Finds the most specific hypothesis matching the training example (hence the name).

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 in h 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

Code:

```
import pandas as pd
import numpy as np
data=pd.read_csv('/content/exp2aiml.csv')
d=np.array(data)[:,:-1]
target=np.array(data)[:,-1]
print(target)
def train(c,t):
    for i, val in enumerate(t):
        if val == "yes":
            specific_hypothesis = c[i].copy()
            break

    for i, val in enumerate(c):
        if t[i] == "yes":
            for x in range(len(specific_hypothesis)):
                if val[x] != specific_hypothesis[x]:
                    specific_hypothesis[x] = '?'
                else:
                    pass
```

```
    return specific_hypothesis
print("n The final hypothesis is:",train(d,target))
```

Output:

```
The final hypothesis is: ['?' '?' '?' 'no' 'natural ' '?']
```

```
print("n The final hypothesis is:",train(d,target))
n The final hypothesis is: ['?' '?' '?' 'no' 'natural ' '?']
```

Interpretation of output:
A person is smart only when his pose is **natural** and he is **not** wearing glasses.

Application:
Q1-What are the Limitations of Find-S Algorithms?
Ans: Few of the limitations of Find-S Algorithm are:

- There is no way to determine if the hypothesis is consistent throughout the data.
- Inconsistent training sets can actually mislead the Find-S algorithm, since it ignores the negative examples.
- Find-S algorithm does not provide a backtracking technique to determine the best possible changes that could be done to improve the resulting hypothesis.

Q2-How many concepts are possible for this instance of space of a given dataset?
Ans: Many algorithms for concept learning organize the search through the hypothesis space by relying on a general-to-specific ordering of hypotheses.

Q3-How many hypotheses can be expressed by the hypothesis language?
Ans: 2, null hypothesis and alternate hypothesis.

Conclusion:

- Successfully implemented the Find-S Algorithm in python using the numpy and pandas libraries.
- Represented the specific hypothesis for the given dataset.
- Determined the output using the specific hypothesis for the given problem statement.