



Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous College Affiliated to University of Mumbai)

BE-ETRX B

Name: Shubham Sawant

Sub- AIML Lab

UID :2019110050

Name of the Experiment:

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: Windows with MATLAB

Data Set Link: [Used-cars-catalog | Kaggle](#)

Dataset Description:

Number of Instances: 19

Number of Attributes: 8

Attribute Information:

1. Transmission – Type of transmission mechanical/automatic.
2. Engine Fuel – Fuel type of the engine.
3. Engine has Gas? – If engine has CNG.
4. Engine Type – gasoline/diesel.
5. Has Warranty? – Whether car has warranty (YES/NO)?.
6. State – New/Owned/Emergency.
7. Drivetrain – Has drivetrain.
8. Is exchangeable? – Whether car can be exchanged (YES/NO)?

Problem Statement: Classify whether car can be exchanged or not given basic details of the car.

Concept: Is exchangeable? YES – Positive example, NO – Negative Example.

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 a_i in h
 If the constraint a_i in h is satisfied by x
 Then do nothing
 Else replace a_i in h by the next more general constraint that is satisfied by x
3. Output hypothesis h



Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous College Affiliated to University of Mumbai)

BE-ETRX B

Name: Shubham Sawant

Code:

Sub- AIML Lab

UID :2019110050

```
File Edit Selection View Go Run Terminal Help • LAB1_AIMLipynb - SEM7 - Visual Studio Code
lab2.py LAB1_AIMLipynb
AIML > EXP1 > LAB1_AIMLipynb > print("The final hypothesis is : ", h)
+ Code + Markdown | Run All | Clear Outputs of All Cells | Go To | Restart | Interrupt | Variables | Outline ... Python 3.10.6 64-bit

import pandas as pd
import numpy as np
df = pd.read_csv(r"C:\Users\Dell\Desktop\Shubham\SEM7\AIML\EXP1\LAB1.csv")
df = df.sample(frac = 0.0005).reset_index(drop=True)

[59] ✓ 0.2s

df.head()

[60] ✓ 0.8s

...

  manufacturer_name  model_name  transmission  color  odometer_value  year_produced  engine_fuel  engine_has_gas  engine_type  engine_capacity  ...  feature_1  feature_2  feature_3  feature_4
0          Citroen         C5         automatic    red        145000         2011        gasoline          False        gasoline         1.6  ...         True         False         True         False
1             Opel        Zafira         mechanical  silver        291000         2001         diesel          False         diesel         2.0  ...         True         False         False         False
2             Rover    400-Series         mechanical   blue        250000         1998        gasoline          False        gasoline         1.4  ...         False         False         False         False
3              Ford         Focus         mechanical  silver        360000         2001         diesel          False         diesel         1.8  ...         True         False         False         False
4              Ford         Probe         mechanical  grey        247000         1996        gasoline          False        gasoline         2.0  ...         True         False         False         True

5 rows x 30 columns

File Edit Selection View Go Run Terminal Help • LAB1_AIMLipynb - SEM7 - Visual Studio Code
lab2.py LAB1_AIMLipynb
AIML > EXP1 > LAB1_AIMLipynb > arr = np.array(df[0:-1])
+ Code + Markdown | Run All | Clear Outputs of All Cells | Go To | Restart | Interrupt | Variables | Outline ... Python 3.10.6 64-bit

arr = np.array(df[0:-1])
target = np.array(df[0:-1])
print("Attributes : ", arr)
print("Target : ",target)

[61] ✓ 0.6s

...
output exceeds the size limit. Open the full output data in a text editor
Attributes : [[['Citroen' 'C5' 'automatic' 'red' 145000 2011 'gasoline' False
'gasoline' 1.6 'sedan' False 'owned' 'front' 9900.0 False
'Минская обл.' 8 1 False True False True False True False False False
True]
['Opel' 'Zafira' 'mechanical' 'silver' 291000 2001 'diesel' False
'diesel' 2.0 'minivan' False 'owned' 'front' 5500.0 False
'Минская обл.' 10 5 False True False False False False False False
False True]
['Rover' '400-Series' 'mechanical' 'blue' 250000 1998 'gasoline' False
'gasoline' 1.4 'hatchback' False 'owned' 'front' 1399.03 False
'Брестская обл.' 1 14 True False False False False False False False
False False]
['Ford' 'Focus' 'mechanical' 'silver' 360000 2001 'diesel' False
'diesel' 1.8 'universal' False 'owned' 'front' 2100.0 True
'Минская обл.' 4 1 False True False False False False False True
True]
['Ford' 'Probe' 'mechanical' 'grey' 247000 1996 'gasoline' False
'gasoline' 2.0 'coupe' False 'owned' 'front' 2400.0 False
'Гомельская обл.' 14 1 False True False False True False False False
False True]
['Ford' 'Sierra' 'mechanical' 'white' 499967 1990 'gasoline' False
'gasoline' 2.3 'universal' False 'owned' 'rear' 600.0 True
'Брестская обл.' 6 5 True False False False False False False False
False False]
['Mercedes-Benz' 'A140' 'mechanical' 'red' 223000 1998 'gasoline' False
```



Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous College Affiliated to University of Mumbai)

BE-ETRX B

Name: Shubham Sawant

Sub- AIML Lab

UID :2019110050

```
file Edit Selection View Go Run Terminal Help
• LAB1_AIMLipynb - SEM7 - Visual Studio Code
lab2.py LAB1_AIMLipynb
AIML > EXP1 > LAB1_AIMLipynb > arr = np.array(d0[:-1])
+ Code + Markdown | Run All | Clear Outputs of All Cells | Go To | Restart | Interrupt | Variables | Outline | Python 3.10.6 64-bit

def training(target,arr):
    for key,value in enumerate(target):
        if value == "YES" :
            h = arr[key].copy()
            break

    for key,value in enumerate(arr) :
        if target[key] == "YES":
            for i in range(len(h)):
                if value[i]!=h[i]:
                    h[i] = '?'
                else :
                    pass

    return h
print("The final hypothesis is :", training(target,arr))

... The final hypothesis is : ['?', '?', '?', 'gasoline', 'No', 'owned', '?']
```

Interpretation of output:

Find-S algorithm moves from specific hypothesis to most general hypothesis of all positive examples. The '?' in the final hypothesis indicates that any value is acceptable for transmission, engine_fuel, engine_has_gas, and drivetrain. From the output we can interpret that cars are generally exchangeable when they have gasoline engine type, don't have warranty and are owned.

Application:

Q1-What are the Limitations of Find-S Algorithms?

Ans: Although FIND-S will find a hypothesis consistent with the training data, it has no way to determine whether it has found the only hypothesis in H consistent with the, or whether there are many other consistent hypotheses as well.

In most practical learning problems, there is some chance that the training examples will contain at least some errors or noise. Such inconsistent sets of training examples can severely mislead FIND-S, given the fact that it ignores negative examples.

Q2-How many concepts are possible for this instance space of a given dataset?

Ans: We have to check distinct instances in X. Therefore, possible instances = $2*3*2*2*1*1*3 = 72$

Q3-How many hypotheses can be expressed by the hypothesis language?

transmission	2
engine_fuel	3
engine_has_gas	2
engine_type	2
has_warranty	1
state	1
drivetrain	3
is_exchangeable	2
dtype: int64	



Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous College Affiliated to University of Mumbai)

BE-ETRX B

Name: Shubham Sawant

Sub- AIML Lab

UID :2019110050

Ans: We have to check distinct hypothesis in X.

Therefore, number of hypotheses expressed = $4*5*4*4*3*3*5*4 = 57,600$

Conclusion:

We can find the most specific hypothesis using Find-S algorithm present in the dataset in positive examples.

We can conclude from the final hypothesis that transmission, engine_fuel, engine_has_gas, and drivetrain can take any values and other attributes are single valued.

We can observe that this algorithm finds just one maximally specific hypothesis though there may be multiple others as well.

The final hypothesis can be used to fit all the positive examples perfectly.