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 **Sub- AIML Lab**

Name: Shubham Sawant **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).

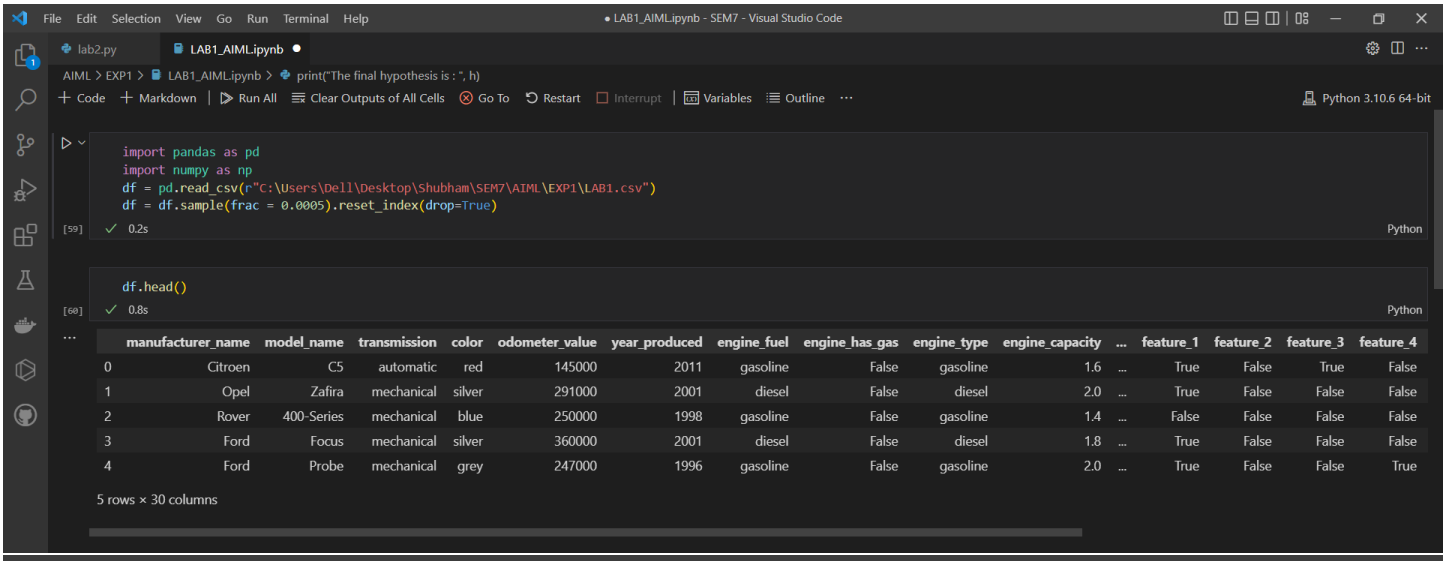
1. Initialize h to the most specific hypothesis in H
2. For each positive training instance x
   For each attribute constraint ai in h
   If the constraint ai in h is satisfied by x
   Then do nothing
   Else replace ai in h by the next more general constraint that is satisfied by x
3. Output hypothesis h

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 **Sub- AIML Lab**

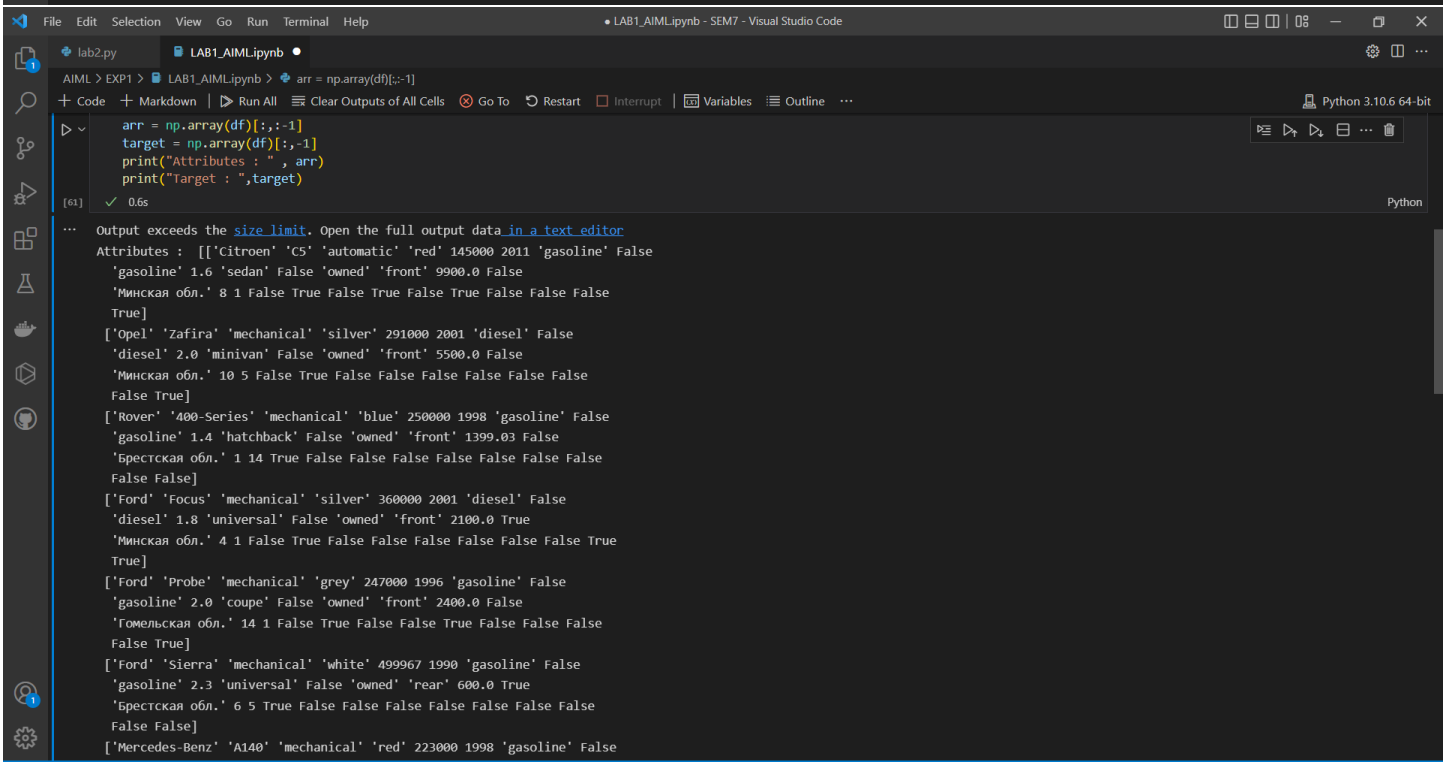
Name: Shubham Sawant **UID :2019110050**

Code:



	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

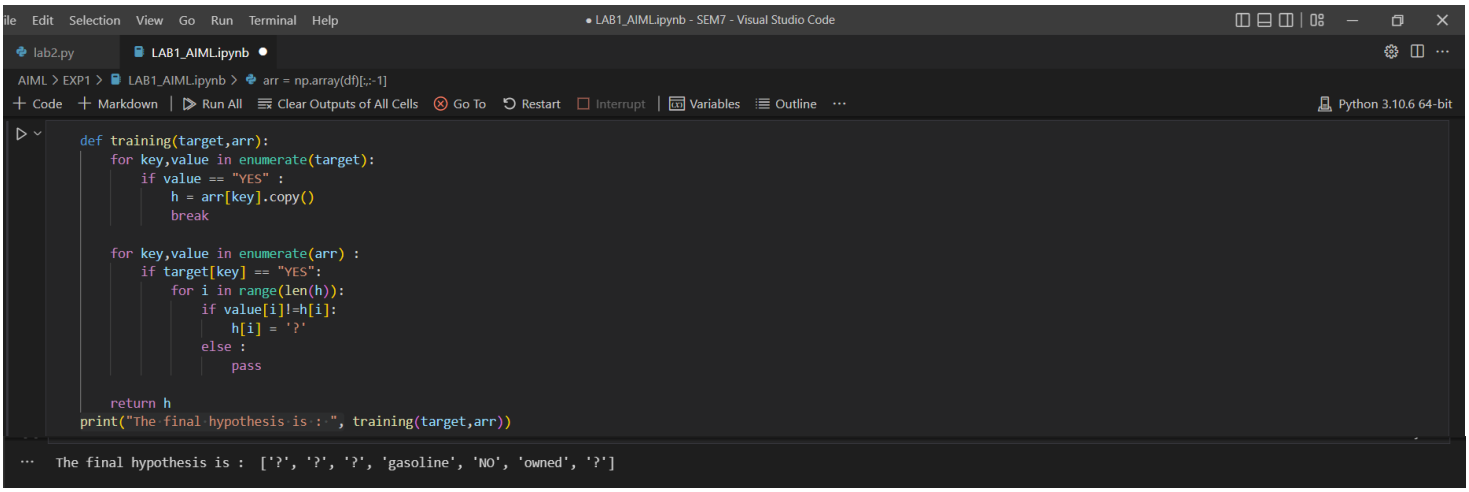
5 rows × 30 columns



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 **Sub- AIML Lab**

Name: Shubham Sawant **UID :2019110050**



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
```

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.