MACHINE LEARNING

ASSIGNMENT 1

Dr. Muhammad Sharjeel

Roha Arslan (FA21-BSE-116)
09/03/2024

COMSATS University Islamabad, Lahore Campus

Course Title:	Machine Learning		Course Code:	CSC354	Credit Hours:	3(3,0)
Resource Person:	Dr. Muhammad Sharjeel		Programme Name:	BSSE		
Semester:	6th	Batch: FA21	Section: A, B, C	Max Marks:	10	

Assignment 1:   **Due Date/Time: Tuesday, 12th March, 23:59**

Submission: Upload the assignment solution to your GitHub account (private repository).

Important instructions: If the assignment requires to submit a python source code (ipython notebook), please write the following information at the start of your file.
 # Date
 # CSC354 – Assignment1 – ML – Concept Learning
 # Your Full Name
 # You Complete Registration Number
 # A brief description of the task

Question1: [CLO-1] - [Bloom Taxonomy Level: <Applying>]
Using Candidate-Elimination algorithm, find (manually) the set of all hypotheses consistent with the following training instances. Show step-by-step complete working of the algorithm. (create and upload the PDF file)

Origin	Manufacturer	Color	Decade	Type	Target
Japan	Honda	Blue	1980	Economy	+
Japan	Toyota	Green	1970	Sports	-
Japan	Toyota	Blue	1990	Economy	+
USA	Chryler	Red	1980	Economy	-
Japan	Honda	White	1980	Economy	+

ANSWER:

ALGORITHM : =

G = maximally general hypotheses in H

S = maximally specific hypotheses in H

For each training example d = <x,c[x]> do :

   If d is a positive example

Remove from G any hypothesis that is inconsistent with d

For each hypothesis s in S that is not consistent with d

Remove s from S

Add to S all minimal generalizations h of s such that

h consistent with d, and some member of G is more general than h

Remove from S any hypothesis that is more general than another hypothesis in S

If d is a negative example

Remove from S any hypothesis that is inconsistent with d

For each hypothesis g in G that is not consistent with d

Remove g from G

Add to G all minimal specializations h of g such that

h consistent with d, and some member of S is more specific than h

Remove from G any hypothesis that is less general than another hypothesis in

$S0 : < \emptyset , \emptyset , \emptyset , \emptyset , \emptyset >$

G0: <?,?,?,?,?>

FOR h1 : <Japan, Honda, Blue, 1980, Economy> +

S1: <Japan, Honda, Blue, 1980, Economy>
G1: <?,?,?,?,?>

FOR h2 : <Japan, Toyota, Green, 1970,Sports> -

S2: <Japan, Honda, Blue, 1980, Economy>
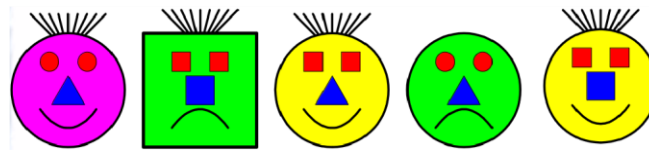G2: {< ?, Honda,?,?,?>, < ?,?, Blue, ?,?>, <?,?,?,1980,?>, <?,?,?,?, Economy>}

FOR h3: <Japan, Toyota, Blue, 1990, Economy> +

S3: <Japan, ?, Blue, ?, Economy>
G3: { < ?,?, Blue, ?,?>, <?,?,?,?, Economy>}

FOR h4: <USA; Chryler, Red, 1980, Economy> -

S4: <Japan, ?, Blue, ?, Economy>
G4: { < ?,?, Blue, ?,?>, <Japan,?,?,?, Economy>}

FOR h5: <Japan, Honda, White, 1980, Economy> +

S5: <Japan, ?,?,?,Economy>
G5: <Japan, ?,?,?,Economy>

S: <Japan, ?,?,?,Economy>

G: <Japan, ?,?,?,Economy>

Question2: [CLO-1] - [Bloom Taxonomy Level: <Applying>]
Using Find-S algorithm, find (manually) a hypothesis that is consistent with the following dataset. Show step-by-step complete working of the algorithm. (create and upload the PDF file)

Note: In the above dataset, the output attribute is 'sad (-)' or 'happy (+)' face. For input attribute(s), you are required to choose all the features depicted in the smiley faces.

ANSWER :

ALGORITHM:
Initialize h to the most specific hypothesis in H
For each positive training instance x
        For each attribute constraint a_i 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
Output hypothesis h

LET h be
        <face shape, eyes shape, hair, nose shape, face colour, smile curve>
Then
        h1 : <circle, circle, yes, triangle, pink, up> +
        h2 : <square, square, yes, square, green, down> -
        h3 : <circle, triangle, yes, triangle, yellow, up> +
        h4 : <circle, triangle, no, triangle, green, down> -
        h5 : <circle, square, yes, square, yellow, up> +

POSITIVE INSTANCES : h1,h3,h5
NEGATIVE INSTANCES : h2,h4

Train/Test Split : 60%/40%
Training set :

        h1 : <circle, circle, yes, triangle, pink, up> +
        h2 : <square, square, yes, square, green, down> -
        h3 : <circle, triangle, yes, triangle, yellow, up> +

Testing set:

        h4 : <circle, triangle, no, triangle, green, down> -
        h5 : <circle, square, yes, square, yellow, up> +

Most specific hypothesis: < ∅ , ∅ , ∅ , ∅ , ∅ , ∅ >
Training:

h0: < ∅, ∅ , ∅ , ∅ , ∅ , ∅ >
h1: <circle, circle, yes, triangle, pink, up>
h3: <circle, ?, yes, triangle, ? , up>

TRAINED MODEL: h3: <circle, ?, yes, triangle, ? , up>

TESTING:

        h4 : <circle, triangle, no, triangle, green, down> - **(predictive output : negative || original output : negative)**
        h5 : <circle, square, yes, square, yellow, up> + **(predictive output : negative || original output : positive)**

ACCURACY = ½ = 50%

If we had implemented Find-S algorithm on the whole data set without splitting
Then our Trained model would have been: h5 <circle, ? , yes, ?, ?, up>

Question3: [CLO-1] - [Bloom Taxonomy Level: <Applying>]
Write a python program (preferably an ipython notebook) to implement Candidate-Elimination algorithm. Give the following training examples as input to the program. The program must output the ordered list of all hypotheses consistent with all the training examples.

Size	Color	Shape	Output
big	red	circle	no
small	red	triangle	no
small	red	circle	yes
big	blue	circle	no
small	blue	circle	yes