

LAB REPORT

**CSE4020 – MACHINE LEARNING LAB**



**(B.Tech. Computer Science Engineering)**

**FALL SEMESTER 2021-2022**

|  |  |
| --- | --- |
| **Name:** | ALOK MATHUR |
| **Reg. No:** | 20BCE2959 |
| **Slot:** | L45+L46 |
| **Faculty Name:** | JAISANKAR N SIR |

**VIT – A Place to Learn; A Chance to Grow**

**ASSIGNMENT 1**

Q1) 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.

***Find – S Algorithm***

1. Initialize h to the most specific hypothesis in H

2. For each positive training instance x

For each attribute constraint a, in h

If the constraint a, is satisfied by x

Then do nothing

Else replace a, in h by the next more general constraint that is satisfied by x

3. Output hypothesis h

***Data Set***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Time** | **Weather** | **Temperature** | **Company** | **Humidity** | **Wind** | **Goes** |
| Morning | Sunny | Warm | Yes | Mild | Strong | Yes |
| Evening | Rainy | Cold | No | Mild | Normal | No |
| Morning | Sunny | Moderate | Yes | Normal | Normal | Yes |
| Evening | Sunny | Cold | Yes | High | Strong | Yes |

***Program***

import numpy as np

import pandas as pd

def training(d,target):

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

if(target[i]=="Yes"):

h1=d[i]

break

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

if(target[i]=="Yes"):

for j in range(0,len(d[0])):

if(d[i][j]!=h1[j]):

h1[j]="?"

else:

continue

print("The value of specific hypothesis for the data set number",i+1," is: ",h1)

print()

print("The best hypothesis is: ",h1)

df=pd.read\_csv('data.csv')

print(“The dataframe is :”,df)

d=np.array(df)[:,:-1]

print("The attributes are :\n",d)

target=np.array(df)[:,-1]

print("The target concepts is : ",target)

training(d,target)

***Output***

The dataframe is :

|  | **Time** | **Weather** | **Temperature** | **Company** | **Humidity** | **Wind** | **Goes** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | Morning | Sunny | Warm | Yes | Mild | Strong | Yes |
| **1** | Evening | Rainy | Cold | No | Mild | Normal | No |
| **2** | Morning | Sunny | Moderate | Yes | Normal | Normal | Yes |
| **3** | Evening | Sunny | Cold | Yes | High | Strong | Yes |

The attributes are :

[['Morning' 'Sunny' 'Warm' 'Yes' 'Mild' 'Strong']

['Evening' 'Rainy' 'Cold' 'No' 'Mild' 'Normal']

['Morning' 'Sunny' 'Moderate' 'Yes' 'Normal' 'Normal']

['Evening' 'Sunny' 'Cold' 'Yes' 'High' 'Strong']]

The target concepts is : ['Yes' 'No' 'Yes' 'Yes']

The value of specific hypothesis for the data set number 1 is: ['Morning' 'Sunny' 'Warm' 'Yes' 'Mild' 'Strong']

The value of specific hypothesis for the data set number 2 is: ['Morning' 'Sunny' 'Warm' 'Yes' 'Mild' 'Strong']

The value of specific hypothesis for the data set number 3 is: ['Morning' 'Sunny' '?' 'Yes' '?' '?']

The value of specific hypothesis for the data set number 4 is: ['?' 'Sunny' '?' 'Yes' '?' '?']

The best hypothesis is: ['?' 'Sunny' '?' 'Yes' '?' '?']