MACHINE LEARNING LAB

July 30, 2020

Manish Singh CS3
Roll No. 17/466

Computer Science and Engineering Department

Rajasthan Technical University, Kota

Contents

1 Problem Statement

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 samples from a .CSV file.(enjoysport.csv)

2 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_i in h
      If the constraint a_i 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

3 Training Set

sky	airtemp	humidity	wind	water	forcast	enjoysport
sunny	warm	normal	strong	warm	same	yes
sunny	warm	high	strong	warm	same	yes
rainy	cold	high	strong	warm	change	no
sunny	warm	high	strong	cool	change	yes

4 Program

```
import pandas as pd

df = pd.read_csv("enjoysport.csv")
print("The data is: \n\n",df)
print("\n"+"*" * 70)
print("The shape of dataset is",df.shape,"\n")
# defining the attributes
num_attributes = 6
#initialize the instance
a = []
#stroing the dataset items in list
for i in range(len(df)):
a.append(df.iloc[i].tolist())

#storing zero at every place
hypo = ['0']* num_attributes
# initialize the specific hypothesis
for j in range(0, num_attributes):
hypo[j] = a[0][j]
print("Most Specific hypotheis is: ",hypo)
# For each positive training instance x
for i in range(len(a)):
if a[i][num_attributes]=='yes':
for j in range(0,num_attributes):# For each attribute constraint ai in h
if a[i][j] != hypo[j]:# Else replace ai in h by the next more general constraint that is satisfied b
hypo[j]="?"
else:# If the constraint ai is satisfied by x
hypo[j]=a[i][j]# Then do nothing
print(" \n\nFor Training instance No:{0} the hypothesis is\n".format(i),hypo)    # Output hypothesis

print("\n The Maximally Specific Hypothesis for a given Training Examples :\n")
print(hypo)
```

5 Output

```
******************** Find S Algorithm ********************
The data is:

      sky airtemp humidity    wind water forcast enjoysport
0  sunny    warm   normal  strong  warm    same         yes
1  sunny    warm     high  strong  warm    same         yes
2  rainy    cold     high  strong  warm  change          no
3  sunny    warm     high  strong  cool  change         yes

*************************************************************************
The shape of dataset is (4, 7)

Most Specific hypotheis is:  ['sunny', 'warm', 'normal', 'strong', 'warm', 'same']


For Training instance No:0 the hypothesis is
 ['sunny', 'warm', 'normal', 'strong', 'warm', 'same']


For Training instance No:1 the hypothesis is
 ['sunny', 'warm', '?', 'strong', 'warm', 'same']


For Training instance No:2 the hypothesis is
 ['sunny', 'warm', '?', 'strong', 'warm', 'same']


For Training instance No:3 the hypothesis is
 ['sunny', 'warm', '?', 'strong', '?', '?']

 The Maximally Specific Hypothesis for a given Training Examples :

['sunny', 'warm', '?', 'strong', '?', '?']
```

6 Result

We have found the most specific hypothesis based on a given set of training data samples using FIND-S algorithm.