



Document ID	2023-24/RPSIT/AI&DS/LM/02	Document Name	Lab Manual
Subject Name	Machine Learning Laboratory	Subject Code	AD3461
Department	Artificial Intelligence & Data Science	Year / Sem	II / IV

Ex. No. 1 :

For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

Aim:

To implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training example stored in a .CSV file.

Algorithm:

Input: Dataset (features and target class)

Output: Specific hypothesis (specific_h), General hypotheses (general_h)

1. Load Dataset from 'Ex1_data.csv'
2. Separate 'concepts' (features) and 'target' (class labels) from Dataset
3. Initialize specific_h with the first instance of concepts
4. Initialize general_h with a list containing a hypothesis as general as possible (all "?")
5. For each instance and its corresponding target in the Dataset:
 - a. If target is "yes" (positive example):
 - i. For each feature in specific_h:
 - If the feature value does not match the instance's feature value, set it to "?"
 - ii. Update each hypothesis in general_h:
 - Set the feature to "?" if it does not match the instance
 - iii. Prune general_h:
 - Keep only those hypotheses that are as general as or more general than specific_h
 - b. If target is "no" (negative example):
 - i. Initialize a temporary list general_h_new
 - ii. For each hypothesis in general_h:
 - For each feature in the hypothesis:
 - If the feature is "?", generate new hypotheses for each possible value except the instance's value
 - If the feature specifies a value different from the instance, copy the hypothesis to general_h_new
 - iii. Update general_h to be general_h_new
 - iv. Prune general_h:
 - Keep only those hypotheses that are at least as general as specific_h

6. Final Pruning of general_h:

a. Remove completely general hypotheses (all "?") from general_h, as they don't contribute to distinguishing instances

7. Return specific_h and general_h as the final specific and general hypotheses consistent with the training data

End Algorithm

Program:

```
1 import numpy as np
2 import pandas as pd
3
4 # Load data
5 data = pd.read_csv('Ex1_data.csv')
6 concepts = np.array(data.iloc[:, :-1])
7 target = np.array(data.iloc[:, -1])
8
9 def candidate_elimination(concepts, target):
10     n_features = concepts.shape[1]
11     specific_h = concepts[0].copy()
12     general_h = [ "?" for _ in range(n_features) ]
13
14     for i, instance in enumerate(concepts):
15         if target[i] == "yes":
16             for x in range(n_features):
17                 # For positive instances, update S and G
18                 if instance[x] != specific_h[x]:
19                     specific_h[x] = '?'
20                 for g in general_h:
21                     g[x] = '?'
22                 general_h = [g for g in general_h if all(feature == '?' or feature == specific_h[x] for x, feature
23 in enumerate(g))]
24             else:
25                 # For a negative instance, refine G
26                 general_h_new = []
27                 for g in general_h:
28                     for x in range(n_features):
29                         if g[x] == "?":
30                             for val in np.unique(concepts[:, x]):
31                                 if instance[x] != val:
32                                     g_new = g.copy()
33                                     g_new[x] = val
34                                     if g_new not in general_h_new:
35                                         general_h_new.append(g_new)
36                                 elif g[x] != instance[x]:
37                                     general_h_new.append(g.copy())
38                 general_h = general_h_new.copy()
39                 general_h = [g for g in general_h if any(all(feature == '?' or feature == s for feature, s in zip(g,
40 specific_h)) for s in specific_h)]
41
42     # Final pruning to remove overly general hypotheses in G
43     general_h = [g for g in general_h if g != [ "?" for _ in range(n_features) ] ]
44     return specific_h, general_h
45
46 s_final, g_final = candidate_elimination(concepts, target)
47
48 print("Final Specific Hypothesis:", s_final)
49 print("Final General Hypotheses:", g_final)
```

Data Set Used (Ex1 data.csv):

Sky,Temp,Humidity,Wind,Water,Forecast,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

Output :

Final Specific Hypothesis: ['sunny' 'warm' '?' 'strong' '?' '?']

Final General Hypotheses: [['sunny', '?', '?', '?', '?', '?'], ['?', 'warm', '?', '?', '?', '?']]