Concept Learning & Evaluation

Exercise 1 : Rule-Based Learning

Given is the following training set D, an excerpt from a wine competition:

	Grape Variety	Bottle-Aged	Acidity	Color Intensity	Award Winning
1	Barbera	yes	mild	high	yes
2	Barbera	yes	strong	high	no
3	Riesling	no	mild	medium	no
4	Barbera	yes	mild	low	yes

Let the set H contain hypotheses that are built from a conjunction of restrictions for attribute-value combinations; e. g. $(\textit{Barbera}, \textit{no}, ?, ?)$.

(a) Apply the Find-S algorithm for the example sequence 1, 2, 3, 4.

Answer

- Initialization: $h_0 : \theta = \ (\bot, \bot, \bot, \bot)$
- Example 1: $h_1 : \theta = \ (\textit{Barbera}, \textit{yes}, \textit{mild}, \textit{high})$
- Example 2: $h_2 = h_1$
- Example 3: $h_3 = h_2$
- Example 4: $h_4 : \theta = \ (\textit{Barbera}, \textit{yes}, \textit{mild}, ?)$

(b) Apply the Candidate-Elimination algorithm for the example sequence 1, 2, 3, 4, and identify the boundary sets H_S and H_G.

Answer

- Initialization. $H_{G_0} = \{(?, ?, ?, ?)\}$; $H_{S_0} = \{(\bot, \bot, \bot, \bot)\}$
- Example 1.
  - $H_{G_1} = H_{G_0}$ (all hypotheses in H_{G_0} are consistent with $\mathbf{x}_1$)
  - $H'_{S_1} = \{(\textit{Barbera}, \textit{yes}, \textit{mild}, \textit{high})\}$ (minimal consistent generalization)
  - $H_{S_1} = H'_{S_1}$ (no $y()$ in H_{S_1} is less specific than another member)
- Example 2.
  - $H_{S_2} = H_{S_1}$ (all hypotheses in H_{S_1} are consistent with $\mathbf{x}_2$)
  - $H'_{G_2} = \{(?, ?, \textit{mild}, ?)\}$ (minimal consistent specializations)
  - $H_{G_2} = H'_{G_2}$ (no $y()$ in H_{G_2} is less general than another member)
- Example 3.
  - $H_{S_3} = H_{S_2}$ (all hypotheses in H_{S_2} are consistent with $\mathbf{x}_3$)
  - $H'_{G_3} = \{(\textit{Barbera}, ?, \textit{mild}, ?), (?, \textit{yes}, \textit{mild}, ?), (?, ?, \textit{mild}, \textit{high})\}$ (minimal consistent specializations)
  - $H_{G_3} = H'_{G_3}$ (no $y()$ in H_{G_3} is less general than another member)
- Example 4.
  - $H_{G_4} = H_{G_3} \setminus \{(?, ?, \textit{mild}, \textit{high})\}$ (not consistent with $\mathbf{x}_4$)

- $H'_{S_4} = \{(\textit{Barbera}, \textit{yes}, \textit{mild}, ?)\}$ (minimal consistent generalization)
- $H_{S_4} = H'_{S_4}$ (no $y()$ in H_{S_4} is less specific than another member)

The final result is $H_S = \{(\textit{Barbera}, \textit{yes}, \textit{mild}, ?)\}$ and $H_G = \{(\textit{Barbera}, ?, \textit{mild}, ?), (?, \textit{yes}, \textit{mild}, ?)\}$.

Exercise 2 : Evaluating Effectiveness

Accuracy is defined as the ratio of correctly classified examples to total examples. Suppose that we are given the following set of six ground-truth examples:

Example	x_1	x_2	c
1	0.75	0.93	1
2	-0.23	1	-1
3	1.20	-0.21	1
4	-0.55	-0.62	-1
5	0.93	0.23	-1
6	0.28	-0.71	-1

Using one of the features x_i ($i \in [1, 2]$), the goal is to learn a classifier $y(\mathbf{x}) = sign(m \cdot x_i)$, where m can be 1 or -1, and i denotes the feature choice.

(a) For a given parameter $m = 1$, calculate the accuracy on the complete dataset for each choice of feature.

Answer

Example	$y(\mathbf{x})$ for $i = 1$	$y(\mathbf{x})$ for $i = 2$	c
1	1	1	1
2	-1	1	-1
3	1	-1	1
4	-1	-1	-1
5	1	1	-1
6	1	-1	-1

$$acc_{i=1} = \frac{4}{6} = 0.67, acc_{i=2} = \frac{3}{6} = 0.5$$

(b) The dataset can be divided into two folds to find the best parametrization for each feature. The model is trained on one fold and evaluated on the other. The first fold includes examples 1,2 and 3; the second fold includes 4, 5, and 6. Compute the accuracy on fold 1 for both parameter choices and feature selections. Do the same for fold 2. Select the best parameter and feature combination for fold 1 and compare to the result on fold two. Interpret the result.

Answer

Example	$y(\mathbf{x})$				c
	$m = 1$		$m = -1$		
	$i = 1$	$i = 2$	$i = 1$	$i = 2$	
1	1	1	-1	-1	1
2	-1	1	1	-1	-1
3	1	-1	-1	1	1
4	-1	-1	1	1	-1
5	1	1	-1	-1	-1
6	1	-1	-1	1	-1

	Accuracy			
	$m = 1$		$m = -1$	
	$i = 1$	$i = 2$	$i = 1$	$i = 2$
Fold 1	1.00	0.33	0.00	0.67
Fold 2	0.33	0.67	0.67	0.33