Concept Learning & Evaluation

Exercise 1 : Concept 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. $\langle \textit{Barbera}, \textit{no}, ?, ? \rangle$.

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

Answer

- Initialization: $\theta = \langle \bot, \bot, \bot, \bot \rangle$
- Example 1: $\theta = \langle \textit{Barbera}, \textit{yes}, \textit{mild}, \textit{high} \rangle$
- Example 2: (ignored)
- Example 3: (ignored)
- Example 4: $\theta = \langle \textit{Barbera}, \textit{yes}, \textit{mild}, ? \rangle$

(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. $H_G = \{\langle ?, ?, ?, ? \rangle\}$; $H_S = \{\langle \bot, \bot, \bot, \bot \rangle\}$
- Example 1. All hypotheses in H_G admit x_1, so H_G remains as-is. Hypothesis $s = \langle \bot, \bot, \bot, \bot \rangle \in H_S$ does not admit x_1 and must be modified:
  - $H_S^+ = \{\langle \textit{Barbera}, \textit{yes}, \textit{mild}, \textit{high} \rangle\}$
  - The hypothesis in H_S^+ has a more general counterpart in H_G, hence $H_S = \{\langle \textit{Barbera}, \textit{yes}, \textit{mild}, \textit{high} \rangle\}$.
  - The only hypothesis in H_S is the most specific one and remains in H_S.
- Example 2. No hypothesis in H_S admits x_2, so H_S remains as-is. Hypothesis $g = \langle ?, ?, ?, ? \rangle \in H_G$ admits x_2 and must be modified:
  - $H_G^- = \{\langle \textit{Riesling}, ?, ?, ? \rangle, \langle ?, \textit{no}, ?, ? \rangle, \langle ?, ?, \textit{mild}, ? \rangle, \langle ?, ?, ?, \textit{medium} \rangle\}$
  - Retaining only those hypotheses in H_G^- with a more specific counterpart in H_S leaves $H_G = \{\langle ?, ?, \textit{mild}, ? \rangle\}$.
  - The only hypothesis in H_G is the most general one and remains in H_G.
- Example 3. No hypothesis in H_S admits x_3, so H_S remains as-is. Hypothesis $g = \langle ?, ?, \textit{mild}, ? \rangle \in H_G$ admits x_3 and must be modified:
  - $H_G^- = \{\langle \textit{Barbera}, ?, \textit{mild}, ? \rangle, \langle ?, \textit{yes}, \textit{mild}, ? \rangle, \langle ?, ?, \textit{mild}, \textit{high} \rangle, \langle ?, ?, \textit{mild}, \textit{low} \rangle\}$

- Retaining only those hypotheses in H_G^- with a more specific counterpart in H_S leaves $G = \{\langle \textit{Barbera}, ?, \textit{mild}, ? \rangle, \langle ?, \textit{yes}, \textit{mild}, ? \rangle, \langle ?, ?, \textit{mild}, \textit{high} \rangle\}$.
- For all three of these, there is no hypothesis in H_G that is more general, hence they remain in H_G.

- Example 4. The hypothesis $\langle ?, ?, \textit{mild}, \textit{high} \rangle \in H_G$ does not admit x_4 and must be deleted. Now, $H_G = \{\langle \textit{Barbera}, ?, \textit{mild}, ? \rangle, \langle ?, \textit{yes}, \textit{mild}, ? \rangle\}$. Hypothesis $s = \langle \textit{Barbera}, \textit{yes}, \textit{mild}, \textit{high} \rangle \in H_S$ must be modified:
  - $H_S^+ = \{\langle \textit{Barbera}, \textit{yes}, \textit{mild}, ? \rangle\}$
  - The hypothesis in H_S^+ has a more general counterpart in H_G, hence $H_S = \{\langle \textit{Barbera}, \textit{yes}, \textit{mild}, ? \rangle\}$.
  - The only hypothesis in H_S is the most specific one and remains in H_S.

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

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:

	p_1	p_2	$c(x_i)$
x_1	0.75	0.93	1
x_2	-0.23	1	-1
x_3	1.20	-0.21	1
x_4	-0.55	-0.62	-1
x_5	0.93	0.23	-1
x_6	0.28	-0.71	-1

Using one of the features p_j ($j \in [1, 2]$), the goal is to learn a classifier $y(x_i) = sign(mx_i)$, where m can be 1 or -1.

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

Answer

	$y(p_1)$	$y(p_2)$	$c(x_i)$
x_1	1	1	1
x_2	-1	1	-1
x_3	1	-1	1
x_4	-1	-1	-1
x_5	1	1	-1
x_6	1	-1	-1

$$acc(p_1) = \frac{4}{6} = 0.67, acc(p_2) = \frac{3}{6} = 0.5$$

(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 x_1, x_2 and x_3, and the second fold includes x_4, x_5 and x_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

	$m = 1$		$m = -1$		$c(x_i)$
	$y(p_1)$	$y(p_2)$	$y(p_1)$	$y(p_2)$	
x_1	1	1	-1	-1	1
x_2	-1	1	1	-1	-1
x_3	1	-1	-1	1	1
x_4	-1	-1	1	1	-1
x_5	1	1	-1	-1	-1
x_6	1	-1	-1	1	-1

	$m = 1$		$m = -1$	
	p_1	p_2	p_1	p_2
Fold 1	1.00	0.33	0.00	0.67
Fold 2	0.33	0.67	0.67	0.33
Average	**0.67**	0.50	0.33	0.50