

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.  $\langle \text{Barbera}, \text{no}, ?, ? \rangle$ .

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

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  $y(x_i) = \text{sign}(mx_i)$ , 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.
- (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.