Lab Class ML:II

Exercise 1 : Concept Learning (Practice)

Given is the following training set D, which you have obtained as co-driver by observing your friend:

	Weekday	Mother-in-the-car	Mood	Time of day	run-a-red-light
1	Monday	no	easygoing	evening	yes
2	Monday	no	annoyed	evening	no
3	Saturday	yes	easygoing	lunchtime	no
4	Monday	no	easygoing	morning	yes

Let the set H contain hypotheses that are built from a conjunction of restrictions for attribute-value combinations; e. g. $\langle$*Monday*, *yes*, ?, ?$\rangle$.

(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 S and G.

(c) What is the version space $V_{H,D}$ for this example?

Exercise 2 : Concept Learning (Background)

(a) Can a version space $V_{H,D}$ contain hypotheses that are neither in the set S nor in the set G? If so, how?

(b) For any two hypotheses $s_1, s_2, s_1 \neq s_2$, from the set S of a version space $V_{H,D}$ holds (check all that apply):

- ☐ $(s_2 \geq_g s_1) \vee (s_1 \geq_g s_2)$
- ☐ $(s_2 \geq_g s_1) \wedge (s_1 \geq_g s_2)$
- ☐ $(s_2 \not\geq_g s_1) \vee (s_1 \not\geq_g s_2)$
- ☐ $(s_2 \not\geq_g s_1) \wedge (s_1 \not\geq_g s_2)$

(c) Which of the two algorithms Find-S and Candidate-Elimination has a stronger inductive bias? Explain your answer.

Exercise 3 : Concept Learning in 2D

Consider the problem of concept learning in the following, rather different, feature space: The set of possible examples is given by all points of the x-y plane with integer coordinates from the interval $[1, 10]$. The hypothesis space is given by the set of all rectangles. A rectangle is defined by the points (x_1, y_1) and (x_2, y_2) (bottom left and upper right corner). Hypotheses are written as $\langle x_1, y_1, x_2, y_2 \rangle$, and assign a point (x, y) to the value 1, if $x_1 \leq x \leq x_2$ and $y_1 \leq y \leq y_2$ hold, with arbitrary, but fixed integer values for x_1, y_1, x_2, y_2 from the interval $[1, 10]$.

Hint: The maximally specific hypothesis s_0 corresponds to a "zero-sized" rectangle that doesn't contain any points with integer coordinates; you may use the symbol $\langle \perp \rangle$.

(a) For the setting described above, formulate the most general hypothesis g_0.

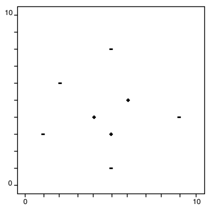
(b) Clarify for yourself how the "more-general" relation $\geq_g$ works in this setting, and check all that apply:

- ☐ $\langle 1,2,3,4\rangle \geq_g \langle 1,1,4,4\rangle$
- ☐ $\langle 2,3,6,7\rangle \geq_g \langle 3,4,5,7\rangle$
- ☐ $\langle 1,1,2,8\rangle \geq_g \langle 1,1,3,3\rangle$
- ☐ $\langle 3,3,9,9\rangle \geq_g \langle 1,1,1,1\rangle$

(c) Given a hypothesis $h = \langle 2,3,5,7\rangle$, and an example $x = (2,7)$ with $c(x) = 0$, determine two hypotheses h_1 and h_2 such that both are minimal specializations of h, and both are consistent with x.

Hint: for the correct answers h_i, there must not exist any hypothesis h' consistent with x where $h \geq_g h'$ and $h' \geq_g h_i$.

(d) Given the following training set:



No.	1	2	3	4	5	6	7	8
Point (x, y)	(5,3)	(9,4)	(1,3)	(5,8)	(4,4)	(5,1)	(6,5)	(2,6)
Class	1	0	0	0	1	0	1	0

Use the Candidate-Elimination algorithm to determine the set of the most general hypotheses G and the set of the most specific hypotheses S. Specify the hypotheses from G and S as $\langle x_1, y_1, x_2, y_2\rangle$ and draw them on the chart.

Hint: pay particular attention, when determining minimal specializations of hypotheses in G, regarding the criteria for keeping the specialized hypotheses.

(e) What happens if an additional example $x_9 = (1,8)$ with $c(x_9) = 1$ is added?

(f) Name a different rule to construct a hypothesis. This rule should have a smaller inductive bias.

Exercise 4 : Evaluating Effectiveness

Consider the following family of classification models:

$$y_\pi(\mathbf{x}) = w \cdot x_\pi$$

where $w \in \{1, -1\}$ is a model parameter learned from data, and $\pi \in \{1, \ldots, p\}$ is a hyperparameter selected manually beforehand. During training, the parameter w is chosen according to the simple learning algorithm shown on the left:

Input: Hyperparameter π and dataset D.

Output: Model Parameter w.

Learn(D, π)

1. **Initialize:** $\mathcal{L}_+ = 0$, $\mathcal{L}_- = 0$
2. **Loop:** For each example $(\mathbf{x}, c(\mathbf{x})) \in D$
   $\mathcal{L}_+ = \mathcal{L}_+ + I(x_\pi, c(\mathbf{x}))$
   $\mathcal{L}_- = \mathcal{L}_- + I(-x_\pi, c(\mathbf{x}))$
3. **If** $\mathcal{L}_+ \leq \mathcal{L}_-$ **Then**
   Return $w = 1$
   Else
   Return $w = -1$

Hint:
The indicator function I is defined as in the lecturenotes slides:

$$I(a, b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{otherwise} \end{cases}$$

Example:
Given
$D = \{((1, -1), 1), ((-1, 1), -1)\}$,
we get:
Learn$(D, 1) = 1$ and
Learn$(D, 2) = -1$

You are given the following dataset D:
(The gray example numbers are only for orientation, and aren't part of the dataset).

Example number		0	1	2	3	4	5	6	7	8	9
$\mathbf{x}$	x_1	1	-1	1	1	1	-1	1	-1	1	1
	x_2	1	1	1	1	1	1	1	-1	1	-1
$c(\mathbf{x})$		1	-1	-1	1	-1	1	-1	-1	1	1

Using the slides from the lecturenotes unit on Evaluating Effectiveness as a guide, complete the following tasks:

(a) Let the hyperparameter π be fixed at $\pi = 1$. Using the algorithm **Learn** given above, train a classifier y_1 on all of D, and determine the training error $Err_{tr}(y_1)$.

(b) Let the examples numbered 7, 8, and 9 in the table now be assigned to the holdout set D_{test}. Leaving $\pi = 1$ as before, train classifier y'_1, and use it to determine the holdout error of y_1.

(c) Using the procedure for model selection with k validation sets, we now want to determine the best possible value for the hyperparameter π. Let $k = 2$, with D_{val_1} containing the examples numbered 0, 1, 2, and 3, and D_{val_2} containing the examples numbered 4, 5, and 6. D_{test} shall once again contain the remaining examples 7, 8, and 9.

Determine the value π^*, and then determine the holdout error for y_{π^*}.