

By Wednesday, 2018-11-15, solutions for the following exercises have to be submitted: 2, 3, 5a-f, 6a-b.

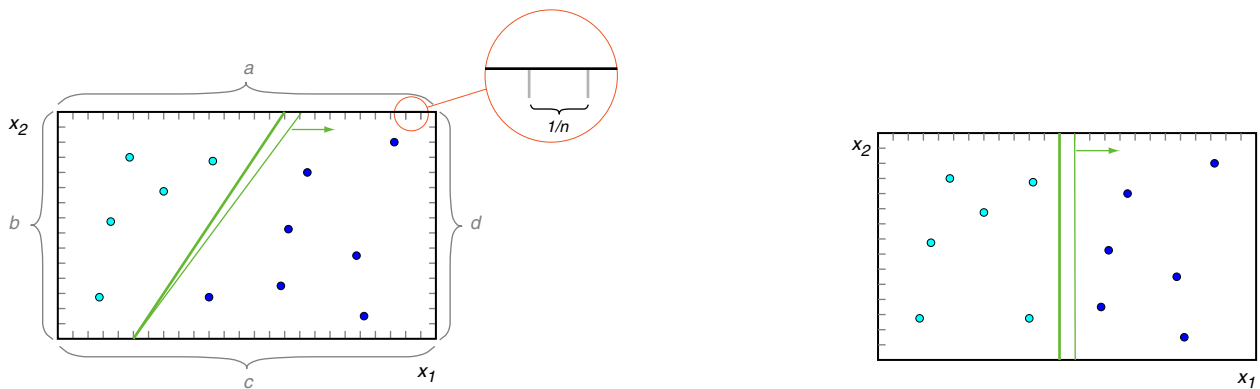
Exercise 1 : Hypothesis Space

Explain the following terms as precisely as possible.

- (a) Hypothesis space
- (b) Version space
- (c) Consistent hypothesis
- (d) Most general hypothesis
- (e) Most specific hypothesis
- (f) Target concept

Exercise 2 : Hypothesis Space

Let x_1 and x_2 be two attributes for the description of objects; the sizes of their domains are $|X_1|$ and $|X_2|$. An object can belong to class 0 or class 1, illustrated as light and dark dots in the figure. Let the hypothesis space H comprise all discrimination lines that can be drawn in the two-dimensional feature space X .



- (a) Compute an upper bound for $|H|$, if arbitrary linear hypotheses are allowed (see left figure).
- (b) Compute an upper bound for $|H|$, if arbitrary linear hypotheses like $x_1 - \beta = 0, \beta \in X_1$ are allowed (see right figure).

Exercise 3 : Concept Learning

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

	Weekday	Mother-in-the-car	Seen-police	Mood	Time of day	Drunk-alcohol	run-a-red-light
1	Monday	no	no	easygoing	morning	nothing	yes
2	Thursday	no	yes	annoyed	evening	something	no
3	Saturday	yes	no	nervous	lunchtime	nothing	no
4	Monday	no	no	easygoing	evening	nothing	yes

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

- (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.

Exercise 4 : Concept Learning (Background)

- (a) The Find-S algorithm considers only positive training examples. Explain whether this property can cause the algorithm to return an inconsistent hypothesis. Assume the hypothesis setup from the lecture.
- (b) Is the hypothesis constructed by the Find-S algorithm dependent on the example order? Explain your answer.
- (c) Which of the two algorithms Find-S and Candidate-Elimination has a stronger inductive bias? Explain your answer.
- (d) Explain the terms "soundness" and "completeness" in the context of the Candidate-Elimination algorithm.

Exercise 5 : P Concept Learning

Develop a simple Python implementation of the Candidate-Elimination algorithm for the learning task discussed in the lecture.

- (a) Hypotheses and examples will be represented as tuples of strings. Use the following functions to generate the minimally and maximally specific hypotheses:

```
def g_0(n):
    return ('?',) * n

def s_0(n):
    return ('T',) * n
```

- (b) Implement a function `more_general(h1, h2)` that returns `True` when $h_1 \geq_g h_2$ holds.
- (c) Implement a function `min_generalization(h, x)` that returns the minimal generalization of hypothesis h that are fulfilled by example x .
- (d) Implement a function `min_specializations(h, domains, x)` for a hypothesis h and an example x . The argument `domains` is a list of lists, in which the i -th sub-list contains the possible values of feature i . The function should return all minimal specializations of h with respect to `domains` which are not fulfilled by x . Example output:

```
>>> min_specializations(h=('?', 'x'),
                        domains=[['a', 'b', 'c'], ['x', 'y']],
                        x=('b', 'x'))
[('a', 'x'), ('c', 'x'), ('?', 'T')]
```

- (e) Implement the Candidate-Elimination algorithm from the [slides](#) as a Python function with the following signature: `candidate_elimination(examples)`, where `examples` is a list of $(n + 1)$ -tuples. The first n elements of each tuple should be the feature values, and the last element is the value of the target concept (you may assume that this value is always either `True` or `False`). Your function should return a 2-tuple containing the sets G and S . Hint: Use the previously implemented functions where appropriate. Start by computing the domains of the features, and generating g_0 and s_0 of the correct dimension.
- (f) Given is the following set of examples for the target concept `EnjoySport`:

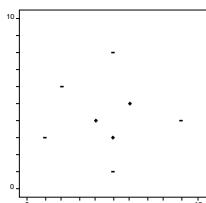
	Sky	Temp	Humid	Wind	Water	Forecast	EnjoySport
1	sunny	warm	normal	strong	warm	same	True
2	sunny	warm	high	strong	warm	same	True
3	rainy	cold	high	strong	warm	change	False
4	sunny	warm	high	strong	cool	change	True

Run your implementation on the examples in the order given above, and report the values of G and S that you obtain.

- (g) Produce a visualization of the hypothesis space spanned by G and S .

Exercise 6 : Concept Learning

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). I. e., hypotheses are functions that 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 $[1, 10]$. 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

- (a) 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. Use at least examples 1, 5 and 8.
- Consider first how to minimally generalize or specify. An example point can be viewed as a rectangle $\langle x, y, x, y \rangle$.
- (b) What happens if an additional example $(1, 8)$ with value 1 is added?
- (c) Name a different rule to construct a hypothesis. This rule should have a smaller inductive bias.
- (d) **P** Modify your implementation from Exercise 5 to handle examples and hypotheses given as rectangles. Run Candidate-Elimination on the eight examples given in the table.