Exercises

Name:

1. **Overfitting vs. Underfitting:**
   What is the problem about "Overfitting" and "Underfitting"? Please explain both situations on the basis of a concrete example.

3. **Detection of Overfitting and Underfitting:**
   How can you detect the occurence of overfitting and underfitting within a concrete machine learning task? Please explain the meaning of partitioning the data in a training, validation and test data partition. How can you detect the following phenomena?

   a. Overfitting has happened.
   b. The space of hypotheses does not allow appropriate modeling of a concrete modeling task.
   c. The data at hand do not allow an appropriate modeling of a certain data modeling task at hand.

4. **Types of Learning Problems**
   Please explain in your own words the characteristics and differences between:

   a. Classification problems.
   b. Regression problems.
   c. Time-series analysis.

   Please explain at least one real-world for each type of learning problem.

   The above mentioned learning problems all belong to the class of supervised learning. What is the difference between supervised and unsupervised learning?

5. **Cross-Validation:**
   Please explain the general functioning of n-fold cross-validation. Why is this strategy meaningful? Is cross-validation suited to evaluate the quality of a model or for evaluating the methodology?

6. **Lattice of hypotheses and consistanca:**
   Given the following training data for a binary classification problem where each example is characterized by 3 boolean attributes a_1, a_2 and a_3. A hypothesis is stated as a triple where '?' stands for an arbitrary value.

a_1	a_2	a_3	Output
0	1	0	1
1	0	0	0
1	1	0	1
1	0	1	0

   a. State the lattice of all hypotheses graphically which can displayed with the more general than relation.

b. Which of the following hypotheses is consistant with the given training data?

[? ? ?]	[ ]
[0 1 0]	[1 1 1]
[? 1 0]	[1 ? 0]
[? ? 0]	[? 1 ?]

7. **Find-S Algorithm:**
Use the training data specified in the previous example in order to perform the Find-S algorithm and show the path in the lattice of hypotheses also stated in the previous example. Does the Find-S algorithm give you the most general or the most specific hypothesis which is consistant with the training data (if such an hypothesis exists)? Please justify your answer.

8. **Candidate Elimination:**
Again considering the training date from example 6 do the following task:

   a. Perform the Candidate-Elimination-Algorithm by hand. In each step please specify S and G in order to end up with the set of consistant hypotheses (i.e. the version space).

   b. Use this result in order to predict the following examples:

a_1	a_2	a_3
0	0	0
1	1	1
0	0	1
0	1	1

   c. Use the Candidate-Elimination-Algorithm with the following additional training data samples and again state S and G in each step of the algorithm. Interpret the results?

a_1	a_2	a_3	Output
1	1	1	1
0	0	0	1

9. **Order of training data in the Candidate Elimination Algorithm:**

Does the order of training data influence the result of the Candidate-Elimination-Algorithm in any way? Can a strategic choice in the ordering of training data help to increase the efficiency of the Candidate-Elimination-Algorithm?

10. **Decision trees for Boolean functions:**

State an appropriate decision tree for each of the following Boolean functions:

- $A \wedge \neg B$
- $A \vee (B \wedge C)$
- $(A \wedge B) \vee (C \wedge D)$

11. **Entropy and Information-gain:**

Assuming the following training instances for a binary classification problem:

a_1	a_2	target
good	small	+
good	small	+
good	large	-
bad	large	+
bad	small	-
bad	small	-

a. Calculate the entropy for the given training data w.r.t. the target.

b. Calculate the information gain of the both attributes a_1 und a_2.

c. Please explain in your own words the meaning of entropy and information gain and why those measures are relevant for decision tree learning.

12. **ID3 Algorithm:**
Assuming the following training instances:

weather	temperature	humidity	wind	water	outlook	nordic walking
sunny	warm	normal	strong	warm	stable	yes
sunny	warm	high	strong	warm	stable	yes
rainy	cool	high	strong	warm	instable	no
sunny	warm	high	strong	cold	instable	yes

a. Perform the ID2-algorithm on the given training date and state the resulting decision tree.

b. Add the following training instance and perform ID3 again in order to learn the corresponding decision tree. Please state also the intermediate steps and report the corresponding entropies and information gains.

weather	temperature	humidity	wind	water	outlook	nordic walking
sunny	warm	normal	light	warm	stable	no

13. **Link between ID3 and Candidate Elimination:**

Consider the case when applying ID3 and Candidate Elimination for a certain learning task. How would you interpret the learned decision tree with respect to S and G of the Candidate Elimination algorithm? Con you give a general statement on this? Please justify your answer.

14. **Pruning:**
Please explain in your own words the meaning of pruning in decision tree learning. What makes the difference between pre- and post-pruning? Would you rather perform pre- or post pruning when applying decision tree learning.

15. **"Lazy Learning" vs. "Eager Learning":**
Case-based reasoning is often called "Lazy Learning" as learning does not happen in the training phase but rather in the application phase when new data samples are to be predicted. In contrast to this "Eager Learning" (like Candidate-Elimination-algorithm, ID3-algorithm, e.g.) generate a model already in the training phase. Please oppose the pros and cons of both approaches and specify concrete machine learning tasks where you think the model based approach or the case based approach is better.

16. **k-Nearest Neighbor classification:**

a. Draft a simple k-nearest neighbor algorithm in order to learn samples with n numerical attributes and a binary target (+/-) in pseudocode. For distance calculation between two samples use the Euclidean distance ($\sqrt{\sum_{i=1}^{n} a_i^2}$).

b. How would such a simple k-Nearest Neighbor algorithm behave if k is equal to the number of training samples? Would such a k be meaningful?

17. **Extensions to the k-Nearest Neighbor Algorithm:**
How could the simple k-nearest neighbor algorithm drafted in 16.a. be extended in a meaningful way.

   a. in order to reduce the required number of training samples?

   b. in order to become more robust w.r.t. noisy training data?