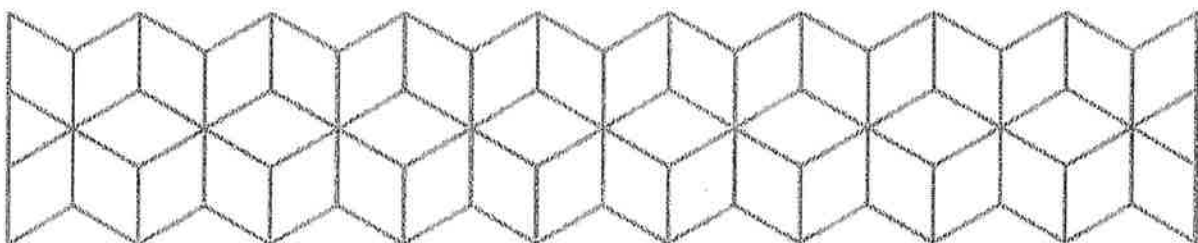


EXAMINATION

| | |
|--|-----------------------------------|
| Course code: ITI43210 | Course: Machin learning |
| Date: May 29 2017 – May 31 2017 | Duration: 3 days |
| Permitted aids: | Lecturer: Roland Olsson |
| The examination: <p>The examination papers consist of 3 pages including this page.</p> <p>Each student must hand in a complete solution to the exam as well as complete project reports even if he or she has been cooperating with other students. Those students who have been working together must state who they have been working with and what is to be regarded as joint work.</p> | |
| Date of announcement of the examination results: June 21 2017 <p>The examination results are available on the Studentweb no later than two workdays after the announcement of the examination results www.hiof.no/studentweb</p> | |



Each student must hand in a complete solution to the exam as well as complete project reports even if he or she has been cooperating with other students. Those students who have been working together must state who they have been working with and what is to be regarded as joint work.

1 The projects (65%)

Attach project reports both on paper and on a CD or DVD disc. The reports should be given both on paper and on .pdf format.

Two complete copies of the reports should be handed in both on paper and on disc.

2 The exam (35%)

Clearly explain how each answer has been produced and give good motivations for all calculations.

A medical dataset contains the following information about patients where the goal is to predict if a patient has hypothyroidism, which is a common disorder of the endocrine system.

```
sex:                male,female.
onThyroxine:        yes,no.
onAntiThyroidMedication:  yes,no.
tumor:              yes,no.
hasHypothyroidism:  yes,no.
```

The training data is as follows.

| sex | onThyroxine | onAntithyroidMedication | tumor | hasHypothyroidism |
|--------|-------------|-------------------------|-------|-------------------|
| female | no | yes | no | yes |
| female | yes | yes | yes | yes |
| male | yes | no | no | yes |
| male | no | yes | no | yes |
| female | no | yes | yes | no |
| female | yes | no | yes | no |
| male | yes | no | yes | no |

The following data is for three new patients that have not yet been diagnosed.

| sex | onThyroxine | onAntithyroidMedication | tumor |
|--------|-------------|-------------------------|-------|
| male | no | no | yes |
| female | yes | no | no |
| male | yes | yes | yes |

- Use entropy heuristics as in the ID3 algorithm to construct a decision tree that perfectly classifies hypothyroidism for the training data. Clearly explain all calculations.
 - Convert the decision tree to rules.
 - How are the three new patients diagnosed by your tree?
- Construct a neural net with perceptron nodes that perfectly classifies the training dataset. Describe how you find the weights in the neural net. You are not supposed to give a general algorithm for training of neural nets nor use any software.
- Assume that a machine learning algorithm gives wrong classifications for 350 out of 1400 test examples. Calculate single and double sided 90% confidence intervals for the error ratio.
- Show how naive Bayes method would classify each of the three new patients. Remember to clearly explain all of your calculations.
- Design a genetic algorithm that learns rules that predict hypothyroidism for a dataset like the one above. Describe and motivate your choice of representation, genetic operators, fitness function and population management as well as other mechanisms you may need to obtain a machine learning method with good generalizing ability.