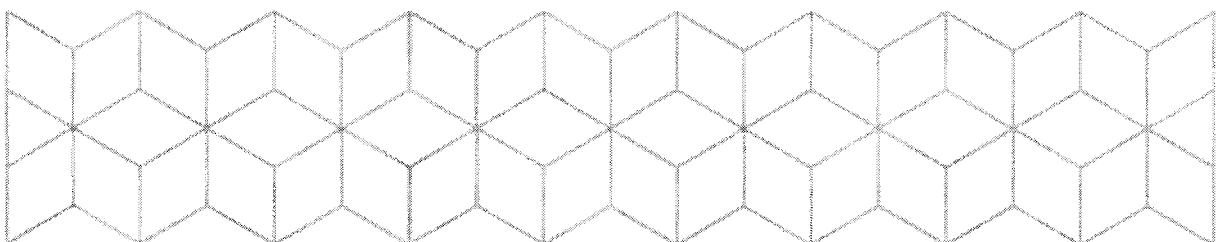


EXAMINATION

Course code: ITI43210	Course: Machine learning
Date/duration 9. May 2016 9.00 11. May 2016 14:00	Lecturer: Jan Roland Olsson
The examination: The examination papers consist of 3 pages inclusive this page. Please check that the examination papers are complete before you start answering the questions.	
Date of announcement of the examination results: 2. June 2016 The examination results are available on the Studentweb no later than two workdays after the announcement of the examination results www.hiof.no/studentweb	



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. The disc should also contain all files that are needed for runs with R, C5.0, neural nets, automatic programming or other machine learning software that has been used.

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

2 The exam (35%)

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

1. A pharmaceutical company has collected genomic data for a number of patients and wish to examine if that data can be used to predict if a person has melanoma. The predictors, that is the inputs, consist of four genetic markers where each marker is either present or absent in a given patient. The dataset is as follows.

Marker1	Marker2	Marker3	Marker4	Melanoma
present	present	absent	absent	sick
absent	absent	present	present	healthy
present	absent	present	absent	sick
present	present	present	absent	healthy
absent	present	absent	present	healthy
absent	present	absent	absent	sick
present	absent	present	present	healthy

- (a) Construct by hand the smallest decision tree that perfectly classifies this dataset.
 - (b) Use Gini impurity heuristics to construct a decision tree that perfectly classifies the dataset. Clearly explain all calculations.
2. The following data set describes whether a given coordinate on a map is on privately or publicly owned land.

(x, y)	Class
(1.5,6)	private
(2,5)	private
(3,4)	public
(3.5,8)	private
(6.5,6.5)	private
(6.5,8)	private
(7.5,6)	public
(8,4)	public
(10.5,5)	public
(10.5,9)	private

Construct a neural net with only one tanh node that perfectly separates private and public property. Carefully illustrate and describe how you find the weights in the neural net. You are not supposed to give a general algorithm for training of neural nets.

- Use another well known machine learning method of your own choice to construct a model for the melanoma dataset. Carefully discuss, explain and illustrate all calculations.
- Use yet another well known machine learning method of your own choice to construct a model for the melanoma dataset. Carefully discuss, explain and illustrate all calculations.