READ THIS FIRST!

1DL073 Natural Computation in Machine Learning, home exam

Allowed help material: Inspera.

Please, answer (in Swedish or English) the following questions to the best of your ability!

This is a home exam, and it is an **individual** exam! You are **not allowed to consult with any other person than me, nor look up things in books or on the Internet**.

For most questions your reasoning is very likely to be more important than the result! So please make sure that you justify your answers to the questions. This is why most of the questions on this exam are 'Essay' questions, to give you full access to the writing tools (including an equation editor) in Inspera. Most answers should not require more than a few sentences though.

The maximum number of points is 40. To get grade 3 (pass) a total of 20 points is required. Grade 4 requires 27 points and grade 5 requires 32 points.

I will be available on Zoom (https://uu-se.zoom.us/j/61117103241) if you have questions, but only **between 9.30 and 10.00**. Zoom will put you in a waiting room where I pick one student at a time. So just wait for your turn.

If you have more questions after that, you may email me (olle.gallmo@it.uu.se). I may not be as quick to respond then though.

Good luck!

Tip: How to write special symbols and Greek letters

How to write special symbols and Greek letters in Inspera

At first glance, it may seem as if some symbols and Greek letters are missing from the equation editor, here in Inspera. But you can insert any Greek letter and many other symbols by writing its name preceeded by a backslash (\), in the equation editor. For example, writing \eta will produce the Greek letter η. If you capitalize the name, the symbol will also be capitalized. For example, \Delta will produce Δ, while \delta will produce δ.

1 Code of Honour 1

Code of Honour

- [ ] I confirm that I will not seek assistance from anyone but the teacher on this exam
- [ ] I confirm that I will not use unauthorized resources to answer the exam questions

Maximum marks: 0

2 ML for noobs

Today is the very first day at university for 90+ fresh computer science bachelor students. They are gathering outside right now, on Campus. Most of them come directly from high school, i.e. without any prior academic background in computer science. How would you explain to them, in a sentence or two, what Machine Learning is? (points for clear, concise, and informative explanations) (2p)

Fill in your answer here

Format

Words: 0

Maximum marks: 2

3 Early Stopping

Many regularization techniques have been discussed on this course. One of them is called *Early Stopping.* Describe how this works and what it tries to achieve! (3p)

Fill in your answer here

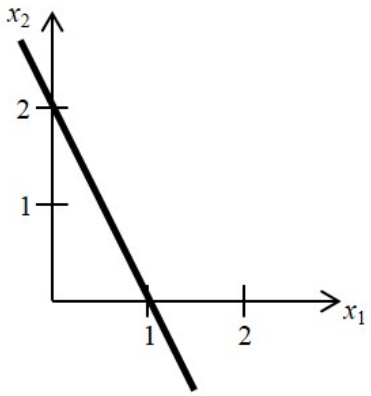
Format

Words: 0

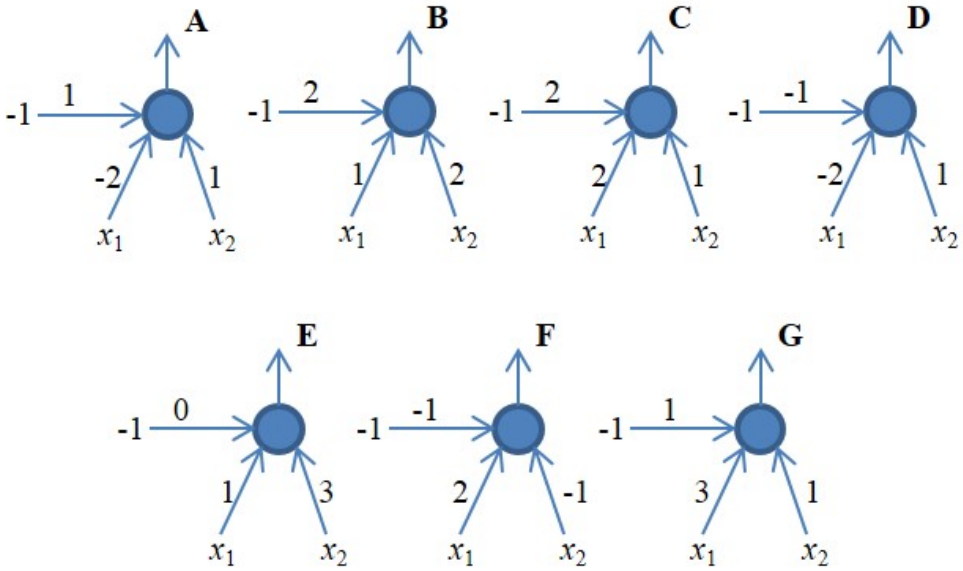
Maximum marks: 3

4 Which perceptron?

A binary perceptron defines a hyperplane in the input space and responds with a 1 on one side of it, 0 on the other. For example this one:



Here are 7 binary perceptrons:



a) One of the perceptrons defines the hyperplane shown above. Which one? (1p)

b) One of the perceptrons defines a hyperplane which crosses origo (0,0). Which one? (1p)

c) Two of the perceptrons define the same hyperplane. They only differ on which side the perceptron would output a 1. Which pair? (1p)

Fill in your answer here

Format

Words: 0

Maximum marks: 3

5 Linear outputs

Neural networks used for function approximation (regression) should have linear outputs. Why? (2p)

Fill in your answer here

Format

Words: 0

Maximum marks: 2

6 MLP & Backprop

Describe a multilayer perceptron (MLP)!

a. Structure: How are the nodes (typically) interconnected? (1p)
b. Forward phase: How does the multilayer perceptron compute its output, given an input vector (x)? (2p)
c. Training: How are the weights adjusted/updated, given an actual output vector (y) and a desired output vector (d), when using the back propagation algorithm? (3p)

Fill in your answer here

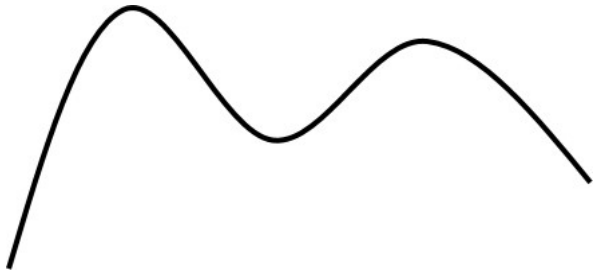
Format

Words: 0

Maximum marks: 6

7 How many hidden nodes?

How many hidden nodes would be required (at least), for a conventional multilayer perceptron (with one hidden layer) to do a good approximation of the function below? Why? (2p)



Fill in your answer here

Format

Words: 0

Maximum marks: 2

8 Radial Basis Functions

The two layers (hidden and output) in a Radial Basis Function network are usually trained separately, and in different ways. The output layer can be trained the usual way (by the delta-rule for example), but how is the hidden layer usually trained? (2p)

Fill in your answer here

Format

Words: 0

Maximum marks: 2

9 Competitive Learning

Describe the Standard Competitive Learning update rule

a) in words! (what is the intention?) (3p)

b) in a formula! (how is this intention achieved?) (1p)

Fill in your answer here

Format

Words: 0

Maximum marks: 4

10 SOFM variance

Consider a classification task where a Self Organizing Feature Map is used to separate two classes. There are two sets of data for each class - one with larger variance than the other (as in the lab you did on this course). What should work best when testing for classification ability - training on the set with low variance and then testing on the set with large variance, or the other way around? (2p)

Fill in your answer here

Format

Words: 0

Maximum marks: 2

11 Q-Learning

A Q-value in Q-Learning can be defined as follows: $Q(s, a)$ is a discounted sum of future rewards, after taking action *a* in state *s*, and assuming that all future actions are greedy.

a) What does it mean to be *greedy* here? (1p)

b) The definition can be written down on a simple recursive form, as an equation in terms of Q itself. How? Write down the equation and explain its parts! (2p)

c) How is this equation then used to define the temporal difference error? (2p)

Fill in your answer here

Format

Σ

Words: 0

Maximum marks: 5

12 Experience Replay

What is *Experience Replay*? Why is it used, and how does it work? (3p)

Fill in your answer here

Format

Words: 0

Maximum marks: 3

13 PSO trade-off

In the Particle Swarm Optimization variant called *lbest*, the particle velocities are updated by a weighted sum of two parts – a trade-off between a *cognitive component* and a *social component*. What are these two components trying to achieve? (what are they striving for?) (2p)

Fill in your answer here

Format

Σ

Words: 0

Maximum marks: 2

14 PSO genotype

What is it, in particle swarm optimization, which most closely corresponds to the *genotype* in a genetic algorithm? (2p)

Fill in your answer here

Format

Words: 0

Maximum marks: 2

15 Population methods v.s. NFL

Particle Swarm Optimization and Genetic Algorithms have in common that they do not depend on gradients of the objective function to work, in contrast to many neural network algorithms. This independence is usually considered an advantage, but the “No Free Lunch” theorem tells us that there must be cases where it is a disadvantage. Describe such a case! (2p)

Fill in your answer here

Format

Words: 0

Maximum marks: 2

16 Code of Honour 2

Code of Honour

- [ ] I confirm that I have not received assistance from anyone but the teacher on this exam
- [ ] I confirm that I have not used unauthorized resources to answer the exam questions

Maximum marks: 0