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 if you have questions, but only during the first hour of the exam. 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 the first hour of the exam, 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

On the previous exam, some students complained that some symbols and Greek letters seem to be missing from the equation editor.

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 Pattern learning

Pattern learning (as opposed to epoch learning) is sometimes called *stochastic*. (For example, stochastic gradient descent, which is a very common special case.) What makes it stochastic?

Fill in your answer here

Format

Words: 0

Maximum marks: 2

3 K-fold cross validation

Explain K-fold cross validation! Why is it used, how does it work, and what is K?

Fill in your answer here

Format

Words: 0

Maximum marks: 3

4 Gaussian XOR

If we take a single binary perceptron, keep the weighted sum but replace the step function (the activation function) with a Gaussian function, would it be able to solve the XOR problem? (just answering yes or no will not give you any points - it's your argument which counts)

Fill in your answer here

Format

Words: 0

Maximum marks: 2

5 What's wrong?

On a previous exam on this course, students were asked to write down the update equations in the Backpropagation algorithm, for a multilayer perceptron with sigmoidal (logistic) nodes. One student answered as follows:

$$\Delta w_{ji} := \eta \delta_j x_i$$

$$\delta_j = \begin{cases} \lambda y_j (1 - y_j)(y_j - d_j) & \text{if } j \text{ is an output node} \\ \lambda y_j (1 - y_j) \sum_k w_{jk} \delta_k & \text{if } j \text{ is a hidden node} \end{cases}$$

There are two errors in these equations. Find and correct them! (you should be able to explain this in words, if you are not comfortable with the equation editor in Inspera)

Fill in your answer here

Format

Words: 0

Maximum marks: 4

6 Close to zero

Backpropagation should work best if we could make sure that the nodes' weighted sums stay close to 0. Why? (There are at least three reasons for this, 1p each)

Fill in your answer here

Format

Words: 0

Maximum marks: 3

7 Network sizing

When training a multilayer neural network on a task, the required number of inputs and outputs depends on how we represent the data. It is more difficult to decide the number of hidden nodes and/or layers.

a) How and why should performance be affected if we have too few hidden nodes? (2p)

b) How and why should performance be affected if we have too many hidden nodes? (2p)

Fill in your answer here

Format

Words: 0

Maximum marks: 4

8 A CNN is a MLP (almost)

A Multilayer perceptron (MLP) consists of two or more layers of non-linear weighted summation units. Defined that way, a convolutional neural network (CNN) is actually also a Multilayer perceptron, except for one particular layer type which is commonly used in CNNs and which are fundamentally different from a MLP layer. Which layer type is that and how is it different?

Fill in your answer here

Format

Words: 0

Maximum marks: 2

9 Voronoi

What is a Voronoi region?

Fill in your answer here

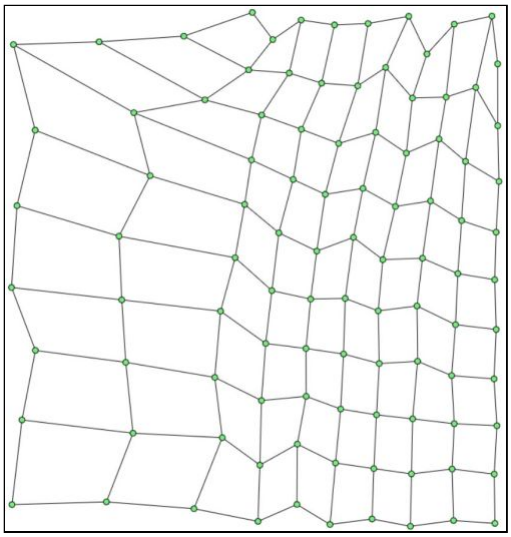
Format

Words: 0

Maximum marks: 2

10 Which algorithm?

An unsupervised neural network has been trained for a while, and the distribution of nodes in the input space now look like this:



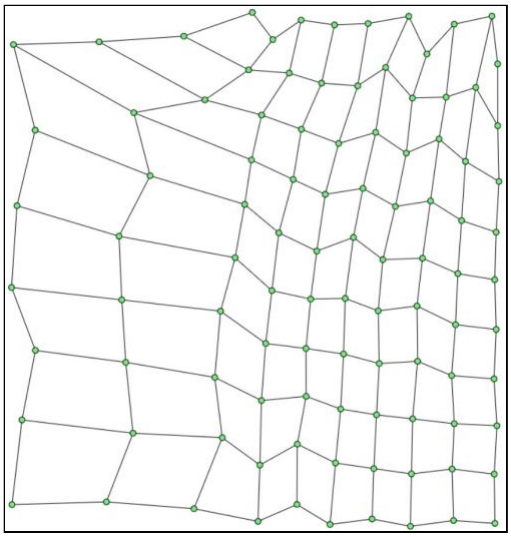
The green circles are nodes (weight vectors) and the lines indicate neighbourhood.
Which of the following unsupervised learning algorithms was most likely used in this experiment?
Select one alternative:

- Self-Organizing Feature Maps
- Hebbian Learning
- Growing Neural Gas
- Competitive Learning
- K-Means

Maximum marks: 1

11 Data deduction

Look at the figure from the previous question again (copied here):



What can you deduce about the distribution of data the network was trained on, given this result?

Fill in your answer here

Format

Words: 0

Maximum marks: 2

12 Non-stationary problems

Which of the following unsupervised learning algorithms should work best for non-stationary input distributions?

Select one alternative:

- Growing Neural Gas
- K-Means
- Competitive learning
- Self-Organizing Feature Maps

Maximum marks: 1

13 True/False about RL

Which of the following statements about reinforcement learning are true?

Select one or more alternatives:

- Q-Learning is greedy, Sarsa explores
- Sarsa is likely to end up with greater Q-values than Q-Learning, if trained on the same problem with the same parameters
- Q-values in Q-learning can only increase over time, not so in Sarsa
- The difference between delayed and immediate reinforcement learning is that rewards are payed/received less often in delayed reinforcement learning
- Reinforcement learning only works if the state machine, which defines the environment, has terminal states
- All previous statements are false

Maximum marks: 1

14 The shortest route

The reward function used in reinforcement learning, should express what we really would like to optimize. It defines the problem. But sometimes it seems it doesn't.

In lab 2 on this course, for example, you trained a robot to find the goal in a simple grid world. The reward function was 0 for all moves, except the very last one, leading directly to the goal. So the problem definition (the reward function) actually does not say that we want to find the *shortest* path. Any path would do!

a) What is it, which makes the algorithm find the shortest path anyway? (2p)

b) If there were no such mechanism, as you described in a), how could we modify the reward function to express the actual goal - to find the shortest path? (2p)

Fill in your answer here

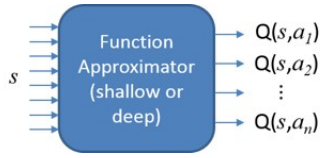
Format

Words: 0

Maximum marks: 4

15 ANNs in Q-Learning

For most real world applications of Q-Learning (or similar reinforcement learning algorithms), the state space is too large for the Q-values to be stored in a table. Therefore, neural networks are often used to estimate the Q-values, like this:



The network has *n* outputs, estimating the Q-values for each of the *n* possible actions in state *s*. Why don't we just compute Q(s,a), i.e. give an action as an extra input and just have one output?

Fill in your answer here

Format

Words: 0

Maximum marks: 2

16 Rank selection

Explain why the use of rank selection, instead of fitness selection, should reduce the risk of premature convergence in an evolutionary algorithm.

Fill in your answer here

Format

Words: 0

Maximum marks: 2

17 GP mutation

Give three typical examples how individuals in genetic programming can be mutated!

Fill in your answer here

Format

Words: 0

Maximum marks: 3

18 Evaporation

The pheromones of real ants evaporate (decay) over time, and so do the pheromones in Ant Colony Optimization (ACO). Why is this evaporation necessary (usually), and how can it be implemented in the pheromone update in ACO?

Fill in your answer here

Format

Words: 0

Maximum marks: 2

19 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