Exam in
Neural Networks and Learning Systems
TBMI26 / 732A55
Home exam - Part II

Date: 2021-03-19
Time: 14.00 - 16.00 (part 1) and **14.00 - 18.00** (part 2)
Teacher: Magnus Borga, Phone: 013-28 67 77

Read the instructions before answering the questions!

Part 2 Consists of four 5-point questions. These questions test deeper understanding and the ability to apply the knowledge to solve problems. All assumptions and calculations made should be presented. Reasonable simplifications may be done in the calculations. **This part needs to be submitted before 18:00**

Students with approved extended examination time (förlängd skrivtid) may divide this extended time between the two parts at their own discretion.

Write your answers by hand and then scan them using a scanner or mobile phone, or write the answers in a separate file using a word processor. The answers may be given in English or Swedish. **If you write by hand, please write clearly using block letters! (Do not use cursive writing.) Answers that are difficult to read, will be dismissed.** The exam should be submitted before the deadline in PDF format. Each part should be handed in as one single PDF file. The PDF files should be named with your LiU-ID followed by a the number of the part of the exam, e.g. 'abcde132-2'.

The maximum sum of points is 20 on each part. To pass the exam (grade 3/C) at least 13 points are required on part 1. For grade 4/B, an additional 10 points on part 2 are required and for grade 5/A, 15 points are required on part 2, in addition to pass part 1.

Note that all forms of collaboration or communication with any person except the course staff is strictly forbidden during the exam!

The result will be reported at 2021-04-13 at the latest.

GOOD LUCK!

AID:	Exam Date: 2021-03-19
Course Code: TBMI26 / 732A55	Exam Code: TEN1

1. Consider a neural network with a single hidden layer, having two inputs, two hidden nodes and two outputs, plus the appropriate biases. The hidden layer uses a ReLU activation function (i.e., $\max(0, x)$), and no activation function is used in the output layer.

   a) Draw said neural network, labeling the relevant elements, and provide mathematical expressions describing the inputs, outputs and operation of each layer. (1p)

   b) Derive the *element-wise* (not matrix form!) weight update rules for online training that would minimize the square error cost function. (2p)

   Our network is now fully trained, but we are curious about its inner workings. We want to discover what kind of inputs strongly activate the various neurons, resulting in large outputs. We can do this using backpropagation. Instead of testing every possible input, we start from some initial input, and progressively update it in a way that would *maximize* some specific output.

   c) Using backpropagation, derive an element-wise update rule for the input that would *maximize* the output of the hidden layer. (1p)

   d) Using backpropagation, derive an element-wise update rule for the input that would *maximize* the output of the network. (1p)

AID:	Exam Date: 2021-03-19
Course Code: TBMI26 / 732A55	Exam Code: TEN1

2. You have the following data:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 2 & 2 & 3 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} -1 & 1 & 1 & 1 & -1 & -1 \end{bmatrix}$$

where $\mathbf{X}$ contains six 2d-samples (one per column), and $\mathbf{Y}$ contains classification labels for the corresponding samples.

a) Perform the first AdaBoost iteration on the data $\mathbf{X}$ with labels $\mathbf{Y}$, using 'decision stumps' as weak classifiers. Sketch the classification problem. (2p)

b) Perform the second AdaBoost iteration on the data, using 'decision stumps' as weak classifiers. Sketch the classification problem. Remember to update the sample weights a second time. (2p)

c) Classify the data with a strong classifier consisting of the two weak classifiers from a) and b). (1p)

Hint: The standard way of updating the weights in the standard AdaBoost method is $d_{t+1}(i) \propto d_t(i)e^{-\alpha_t y_i h_t(\mathbf{x})}$, where $\alpha_t = \frac{1}{2} \ln \frac{1-\epsilon_t}{\epsilon_t}$.
Hint 2: The final strong classifier: $H(\mathbf{X}) = sign(\sum_{t=1}^{T} \alpha_t h_t(x))$

AID:	Exam Date: 2021-03-19
Course Code: TBMI26 / 732A55	Exam Code: TEN1

3. a) We have a non-linear feature mapping $\Phi(\mathbf{x})$ such that

$$\Phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ 1 \end{pmatrix}$$

Show how the kernel $K(\mathbf{x}, \mathbf{y})$ associated with this feature mapping can be expressed as a non-linear function of scalar product in the input space. (2p)

b) Using the kernel function K you derived in a) and an N-dimensional input space, i.e. $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$, calculate the number of dimensions in the mapped feature space. (3p)

AID:	Exam Date: 2021-03-19
Course Code: TBMI26 / 732A55	Exam Code: TEN1

4. The figure shows a deterministic state model with the corresponding reward function. The states are numbered 1 to 6 and the arrows represent the actions "down" and "right". The digits next to the arrows denote reward.

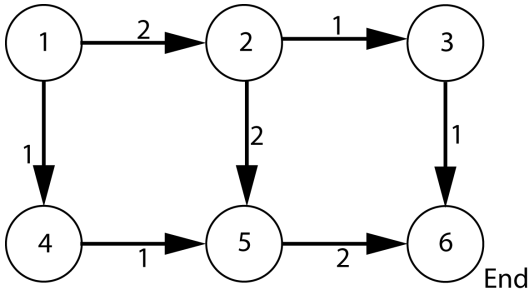

Figure 1: State model.

a) Derive the optimal Q- and V-functions of the system as functions of the discount factor γ $(0 < \gamma < 1)$. (2p)

b) Calculate the Q-function after 7 steps of Q-learning. Start with the Q-function $Q(s, a) = 0$ for every state and action. Use the discount factor γ, learning rate α and apply Q-learning using the following actions:

$3 \to 6, \quad 5 \to 6, \quad 2 \to 3, \quad 2 \to 5, \quad 4 \to 5, \quad 1 \to 2, \quad 1 \to 4.$

Please note that it is possible to perform Q-learning even though these actions do not follow a connected path through the state graph. (3p)