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

Date: 2020-06-10
Time: 14.00 - 16.00 (part 1) and **15.30 - 18.00** (part 2)
Teacher: Magnus Borga, Phone: 013-286777

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**

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.

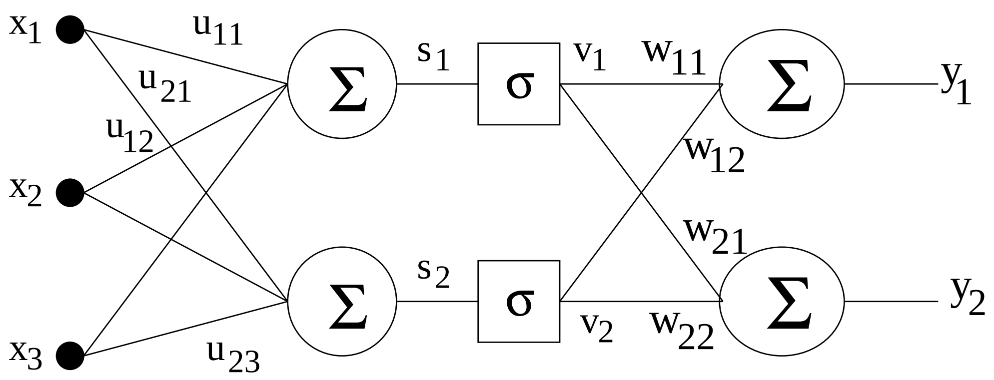
Note that all forms of collaboration or communication with any person except the course staff is strictly forbidden during the exam!

The result will be reported at 2020-07-01 at the latest. The exams will then be available at "studerandeexpeditionen" at IMT.

GOOD LUCK!

AID:	Exam Date: 2020-06-10
Course Code: TBMI26 / 732A55	Exam Code: TEN1

1. Consider the following neural network.



tanh is used as activation function.

a) Derive an expression for the weight updates of both the layers. (3p)

b) Of what use is the activation function in the hidden layer? (1p)

c) Explain using words how the net is trained in practice and how it is used for classification. (1p)

AID:	Exam Date: 2020-06-10
Course Code: TBMI26 / 732A55	Exam Code: TEN1

2. The correlation of a 1D feature map channel $f_i(x)$ with a kernel $h_{ij}(x)$ of size $K = 2L+1$ is defined

$$g_{ij}(x) = (h_{ij} \square f_i)(x) = \sum_{\alpha=-L}^{L} h_{ij}(\alpha) f_i(x+\alpha) \ .$$

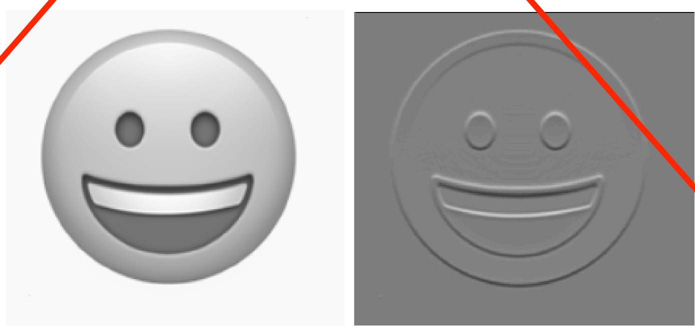
We now look at the example of a two-channel input feature map ($i = 0, 1$) and a two-channel output feature map ($j = 0, 1$), i.e., we use four kernels in total. The final output is computed as $y_j(x) = \sum_i g_{ij}(x)$.

a) Perform the four correlations below. All values outside the feature map channels f_i are equal to zero. In the arrays f_i and h_{ij}, the respective number written in bold face is at position $x = 0$. Note that g_{ij} is only a part of the correlation result (*'same'*). (2p)

h_{ij}	$\square$	f_i	=	g_{ij}	
1 **2** 1	$\square$	**0** 1 0 1	=	**?** ? ? ?	$i = 0, j = 0$
0 **2** 3	$\square$	**0** 1 0 1	=	**?** ? ? ?	$i = 0, j = 1$
2 **1** 0	$\square$	**2** 0 1 0	=	**?** ? ? ?	$i = 1, j = 0$
1 **0** 1	$\square$	**2** 0 1 0	=	**?** ? ? ?	$i = 1, j = 1$

b) We now assume 5 input channels, 2 output channels, and 2-dimensional kernels of size 3×3.
How many parameters need to be learned (no bias coefficient)? (1p)

c) Now we consider 2D correlation. Consider the two images below.



Write a 3×3 correlation kernel that produces the right image from the left one. (2p)

AID:	Exam Date: 2020-06-10
Course Code: TBMI26 / 732A55	Exam Code: TEN1

3. You have the following data:

$$\mathbf{X} = \begin{bmatrix} -1 & -1 & 1 & 0 & 0 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix}$$

where $\mathbf{X}$ contains six samples with three feature attributes, and $\mathbf{Y}$ contains classification labels for the corresponding samples.

a) Perform the first AdaBoost iteration on the data $\mathbf{X}$ using the labels $\mathbf{Y}$ using brute force optimization and 'decision stumps' as weak classifiers. (4p)

b) After how many iterations will the training accuracy reach 100%? (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 $\alpha_t = \frac{1}{2}\ln\frac{1-\epsilon_t}{\epsilon_t}$.

AID:	Exam Date: 2020-06-10
Course Code: TBMI26 / 732A55	Exam Code: TEN1

4. The figure shows two different deterministic state models and the corresponding reward functions. The states are enumerated and arrows represent actions. The numbers close the actions denotes the corresponding rewards (to be maximized). If the system reaches a state denoted “End”, no more reward is obtained.

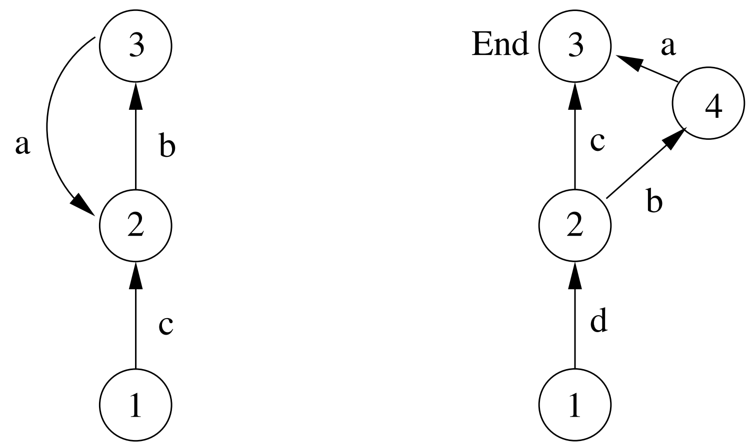


Figure 1: State models A and B.

a) Calculate the optimal Q and V functions for system A as a function of $0 < \gamma < 1$ where (a,b,c) = (1,2,1). (2p)

b) Calculate the optimal Q and V functions for system B as a function of $0 < \gamma < 1$ where (a,b,c,d) = (3,1,2,1). (3p)