Exam in
Neural Networks and Learning Systems
TBMI26 / 732A55

Time:	2023-06-08, 14-18
Teacher:	Magnus Borga, Phone: 013-28 67 77
Allowed additional material:	Calculator, English dictionary

Read the instructions before answering the questions!

The exam consists of two parts:

Part 1 Consists of 10 one-point and 5 two-point questions. The questions test general knowledge and basic understanding of central concepts in the course. The answers should be short and given **on the blank space after each question or in the indicated figure**.

Part 2 Consists of 4 five-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. **All calculations and answers on part 2 should be on separate papers! Do not answer more than one question on each paper!**

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.

The result will be reported at 2023-06-29 at the latest. The exams will then be available at studerandeexpeditionen at IMT.

GOOD LUCK!

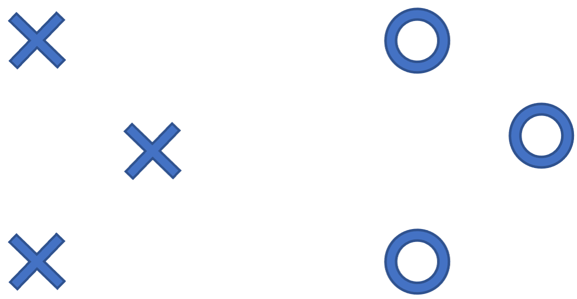
AID:	Exam Date: 2023-06-08
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 1 (Write your answers in the exam handout, directly after each question!)

One-point questions

1. Machine learning can be divided into supervised, unsupervised and reinforcement learning. Mention one example of learning algorithms for each of these three classes.

2. Draw the line for which the discriminant function $f(\mathbf{x}) = 0$ for a linear SVM without slack variables and mark the support vectors in the figure below



3. Outliers, such as mis-labeled data, can cause problems for the AdaBoost algorithm. How can this problem be detected?

AID:	Exam Date: 2023-06-08
Course Code: TBMI26 / 732A55	Exam Code: TEN1

4. What is the purpose of the hidden layers in a multi-layer perceptron classifier?

5. How can the problem of overfitting be avoided when training a neural network?

6. Mention a classifier that does not use a parameterized discriminant function.

7. How can you increase the generalization performance of a classifier when you have a limited training data set?

8. Suppose that you know the Q-function values for a certain state. How do you determine the V-value for that state?

9. Konsider the kernel function

$$k(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{3} e^{x_i + y_i}$$

Write the corresponding non-linear mapping of the input feature vector, i.e. $\varphi(\mathbf{x})$.

10. Assume you are evaluating a binary classifier (i.e. two classes) and there are twice as many samples in the second class compared to the first class. The classifier classifies 80% and 60% correct from class 1 and 2 respectively. What is the accuracy of this classifier?

AID:	Exam Date: 2023-06-08
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Two-point questions

11. What kind of tasks do you typically use the ”U-Net” for? (1p) Draw this architecture schematically! (1p)

12. In deep neural networks, the ReLU activation function is commonly used instead of a sigmoid activation function. Give a brief explanation for why this is the case (1p) and draw both activation functions (1p).

13. Mention two advantages with convolutional neural networks compared to a simple fully connected network.

14. What kind of optimization method is used when trining a decision stump in AdaBoost? Why is this method chosen?

15. Explain the difference between using paired vs. unpaired training for generative adversarial networks. (1p) What is the main advantage and disadvantage with each method? (1p)

AID:	Exam Date: 2023-06-08
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 2 (Write your answers on separate papers. Not more than one answer on each paper!)

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 *tanh* activation function 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)

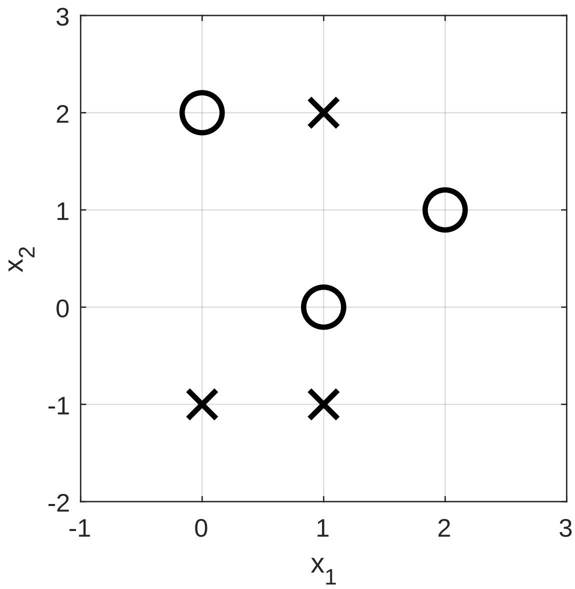
   Our network is now fully trained, but we are curious about its inner workings. We want to discover what kind of inputs strongly activate the various neurons, resulting in large outputs. We can do this using backpropagation. Instead of testing every possible input, we start from some initial input, and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

   Hint: $\frac{\partial}{\partial x} tanh(x) = 1 - tanh^2(x)$

2. You have the following data:



where "circles" is the positive class and "crosses" is the negative class, i.e.

$$\mathbf{X} = \begin{bmatrix} 0 & 1 & 2 & 0 & 1 & 1 \\ 2 & 0 & 1 & -1 & -1 & 2 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix}$$

We are using AdaBoost on this data by brute force optimization of 'decision stumps' as weak classifiers.

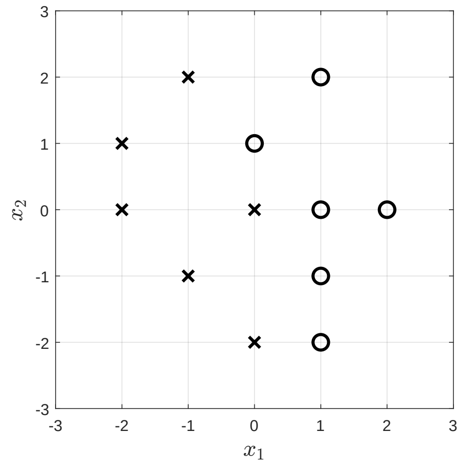
a) Perform the first two iterations of AdaBoost on this data. Calculate the updated weights for both iterations. (4p)
Hint: Using exact values in the calculations (i.e. no approximations with calculator) will give very clean answers.

b) What is the accuracy of the strong classifier built by the two weak classifiers in a)? (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 $\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$.

AID:	Exam Date: 2023-06-08
Course Code: TBMI26 / 732A55	Exam Code: TEN1

3. The data points in the figure have two features (x_1 and x_2) and belong to either the class *crosses* or the class *circles*:



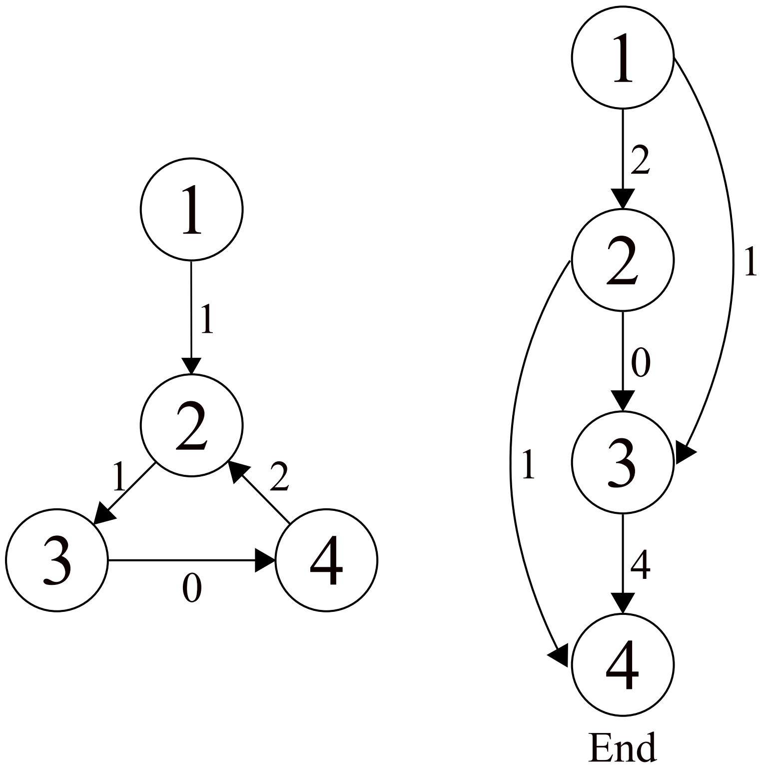
Perform Linear Discriminant Analysis (LDA) on the data to reduce the dimensionality to one dimension that separates the two classes optimally. Calculate the reduced data. (5p)

Hint 1: The inverse of a 2×2-matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Hint 2: If you use fractions and roots instead of rounded values all the way, the final answer will be very clean.

AID:	Exam Date: 2023-06-08
Course Code: TBMI26 / 732A55	Exam Code: TEN1

4. The figure shows two different deterministic state models and the corresponding rewards. The states are enumerated and the arrows represent actions. The numbers close to the arrows show the corresponding rewards. If the system reaches a state denoted "End" no additional rewards are given, i.e. the V-function is defined as 0 in such a state. An optimal policy is in this context the policy which maximizes the reward.



Figur 1: The state models A (left) and B (right).

a) Calculate the optimal Q- and V-functions for system A as functions of $0 < \gamma < 1$. (2p)

b) Calculate the optimal Q- and V-functions for system B as functions of $0 < \gamma < 1$. (3p)