Exam in
Neural Networks and Learning Systems
TBMI26 / 732A55

Time:	2023-08-26, 8-12
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 2023-09-15 at the latest. The exams will then be available at studerandeexpeditionen at IMT.

GOOD LUCK!

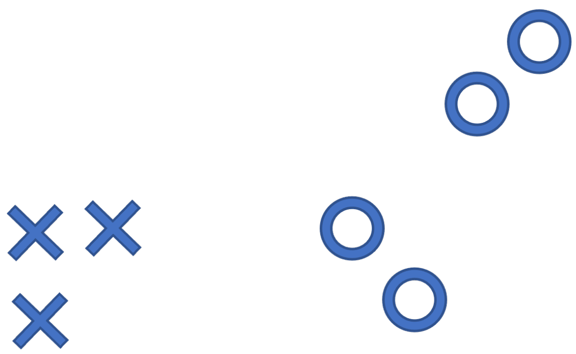
AID:	Exam Date: 2023-08-26
Course Code: TBMI26 / 732A55	Exam Code: TEN1

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

One-point questions

1. Machine learning is often divided into the categories Supervised, Unsupervised and Reinforcement learning. Categorize the following three learning methods accordingly:
   - Q-learning
   - Support vector machines
   - Principal component analysis

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



3. After you have trained an AdaBoost classifier you note that one training sample has received a larger weight than other training samples. What could be the reason and effect of this?

4. Assume we want to use a simple one-layer perceptron to predict which persons pass a certain physical test, based on the persons' height and weight. How many parameters do we have to train?

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

6. How should you change the value of k in k-NN if you want to decrease the risk of overfitting?

7. Assume you have a set of images and want to train a convolutional neural network to classify images. How can you pre-process the training data in order to avoid over-fitting?

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

AID:	Exam Date: 2023-08-26
Course Code: TBMI26 / 732A55	Exam Code: TEN1

9. Write the general definition of a kernel function $k(\mathbf{x}, \mathbf{y})$ in terms of a nonlinear mapping $\Phi(\mathbf{x})$.

10. You evaluate a classifier you have just trained and find the following *confusion matrix*. What is the classification accuracy?

		Predicted class		
		Class 1	Class 2.	Class 3
Actual class	Class 1	100	20	10
	Class 2.	10	50	30
	Class 3	10	20	150

AID:	Exam Date: 2023-08-26
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Two-point questions

11. If you want to solve an image segmentation task with deep learning, what kind of network architecture would you use? (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.

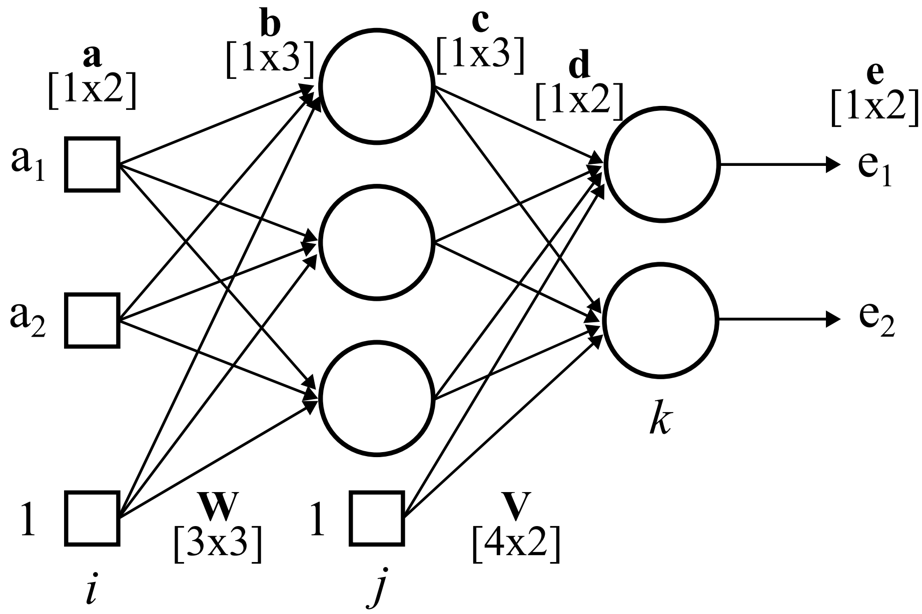
14. Write down the *empirical risk/0-1* loss function for training a classifier (1p). How can a classifier be trained using this loss function (1p)?

15. What are the two main building blocks called in Generative Adversarial Networks? (1p) Briefly explain the tasks of these two building blocks (1p)

AID:	Exam Date: 2023-08-26
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 the following neural network:



We want to use it to classify some data. However, we are not getting very good results, because we forgot to include any activation functions, that is, $\sigma(x) = x$.

a) Derive the *element-wise* weight update rules for online training that would minimize the square error cost function for this network. (2p)

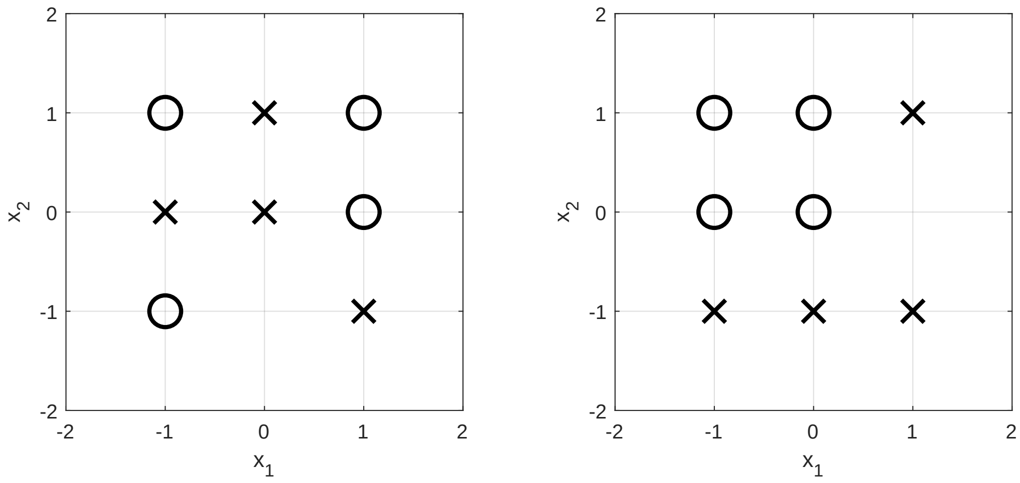
We noticed our mistake, and decided to add tanh activation functions to both the hidden and the output layers.

b) Derive the *element-wise* weight update rules for the corrected network. (2p)

c) Describe the differences in the types of decision boundaries that can be learned by each network. (1p)

AID:	Exam Date: 2023-08-26
Course Code: TBMI26 / 732A55	Exam Code: TEN1

2. You have the following data:



(a) Dataset A (b) Dataset B

where "circles" is the positive class and "crosses" is the negative class.

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

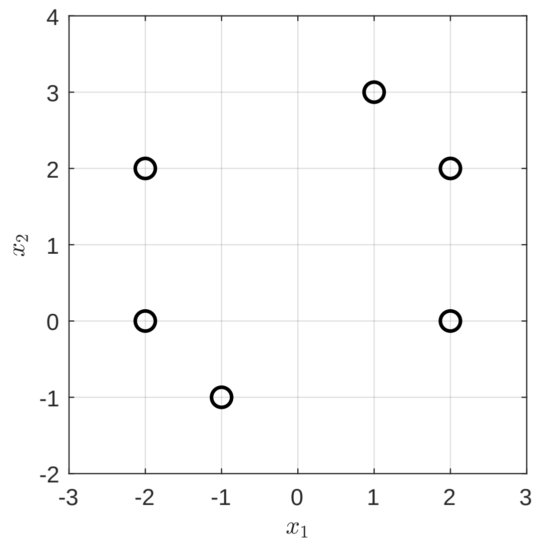
a) Write all valid choices of optimal weak classifier for the first iteration of AdaBoost on dataset A. Weak classifiers that give the same classification are considered to be the same in this case. (2p)

b) Perform the first two iterations of AdaBoost on dataset B. Calculate the updated weights for both iterations. (3p)
Hint: Using exact values in the calculations (i.e. no approximations with calculator) will give very clean answers.

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: 2023-08-26
Course Code: TBMI26 / 732A55	Exam Code: TEN1

3. We have the data



defined by the matrix

$$\mathbf{X} = \begin{bmatrix} -2 & -2 & -1 & 2 & 1 & 2 \\ 0 & 2 & -1 & 0 & 3 & 2 \end{bmatrix}$$

a) We think the data volume is too large. Use principal component analysis to reduce the data to one dimension. (3p)

b) How much of the information (total variance) in the data is accounted for by the first principle direction? (1p)

c) Show how the original signal can be approximately reconstructed given the results from a), and perform the calculations. (1p)

Hint: The eigenvalues of a 2×2-matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ are:

$$\lambda_1 = \frac{T}{2} + \sqrt{\frac{T^2}{4} - D} \text{ and } \lambda_2 = \frac{T}{2} - \sqrt{\frac{T^2}{4} - D},$$

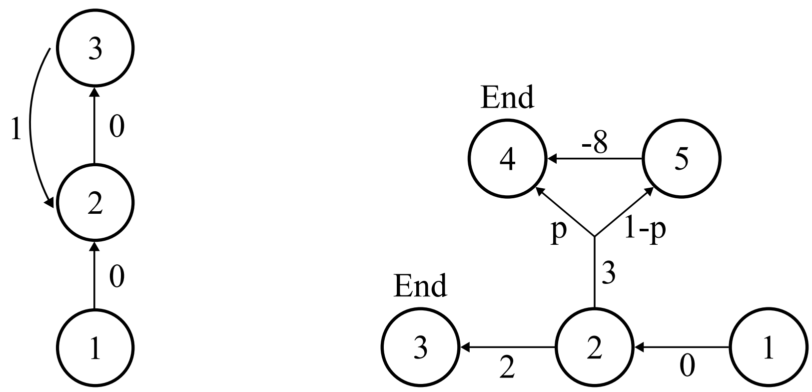
with trace $T = a + d$ and determinant $D = ad - bc$. The corresponding eigenvectors are

$$\mathbf{e}_1 = \begin{pmatrix} \lambda_1 - d \\ c \end{pmatrix} \text{ and } \mathbf{e}_2 = \begin{pmatrix} \lambda_2 - d \\ c \end{pmatrix}$$

Hint: Using exact values in the calculations (i.e. no approximations with calculator) will give very clean answers.

AID:	Exam Date: 2023-08-26
Course Code: TBMI26 / 732A55	Exam Code: TEN1

4. The figure below shows two different state models and the corresponding reward function. The states are numbered and the possible actions are represented by arrows. The number next to the arrow designates the reward for taking that action. If the system has reached an "End" state no more rewards are given.



Figur 2: State models A and B.

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

b) In system B, the action to go up in state 2 will result in the state randomly changing to state 4 or 5, where the arrows p and 1-p indicate the probability of the respective path. Regardless of the random state change the reward for the action is 3. Calculate the optimal Q- and V-functions for system B as a function of $0 < \gamma < 1$ and $0 < p < 1$. (2p)

c) In system B, given $\gamma = 0.5$, for what values of p will the path $1 \rightarrow 2 \rightarrow 3$ be the optimal policy from state 1? (1p)