Exam in
Neural Networks and Learning Systems
TBMI26 / 732A55

Time:	2023-03-20, 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 2023-04-12 at the latest. The exams will then be available at studerandeexpeditionen at IMT.

GOOD LUCK!

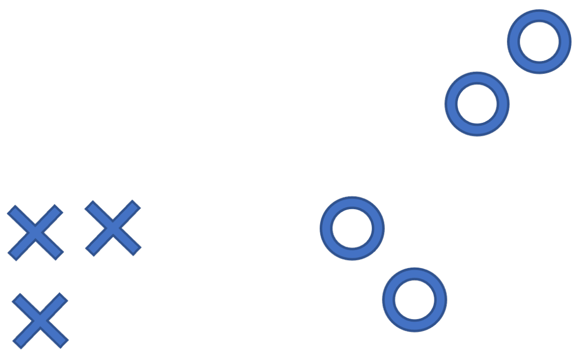
AID:	Exam Date: 2023-03-20
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 1

One-point questions

1. Machine learning is often divided into the categories Supervised, Unsupervised and Reinforcement learning. Categorize the following three learning methods accordingly:
   - k-means
   - Q-learning
   - Ada-Boost

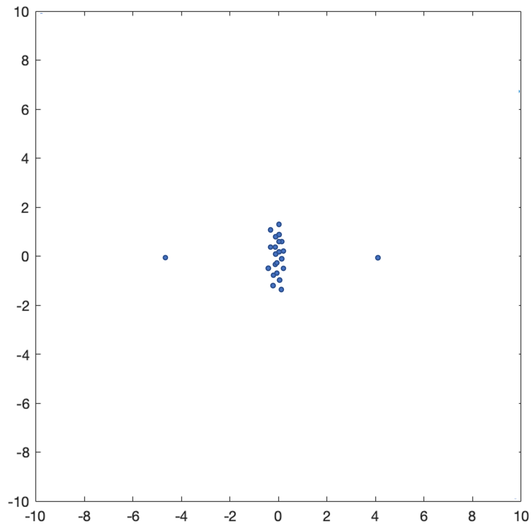
2. Draw the line for which the discriminant function $f(\mathbf{x}) = 0$ for a linear SVN without slack variables and mark the support vectors in the figure below



3. Why can outliers in the training data be a problem in the discrete AdaBoost algorithm?

AID:	Exam Date: 2023-03-20
Course Code: TBMI26 / 732A55	Exam Code: TEN1

4. Draw the first principal component of the distribution in the figure below!



5. When training a complex model such as a multi-layer neural network, there is a risk for *overfitting*. What is the consequence of overfitting when you are then using the trained model?

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

AID:	Exam Date: 2023-03-20
Course Code: TBMI26 / 732A55	Exam Code: TEN1

7. What is the purpose of data augmentation in the context of training a supervised learning system?

8. A system is trained with reinforcement learning. What will be the difference in behavior be if the discount factor is set to a large value compared to a low value?

9. Consider the following non-linear mapping from the N-dimensional input space to a non-linear feature space:

$$\varphi(\mathbf{x}) = \begin{pmatrix} e^{-x_1^2} \\ e^{-x_2^2} \\ \vdots \\ e^{-x_N^2} \end{pmatrix}$$

Write the corresponding kernel function as a sum of exponentials, i.e., in this form:

$$k(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{?} e^{??}$$

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

		Classified label	
		Class 1	Class 2
Actual label	Class 1	80	20
	Class 2	30	90

AID:	Exam Date: 2023-03-20
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. The sigmoid class of activation functions (e.g., *tanh*) does not work well in the hidden layers in a deep neural network. Briefly explain why (1p) and suggest an activation function that solves this problem (1p).

13. Write a 3-by-3 convolution kernel that detects horizontal edges (i.e., transitions from dark to bright) in an image. The kernel should not be sensitive to the mean intensity of the image neighbourhood.

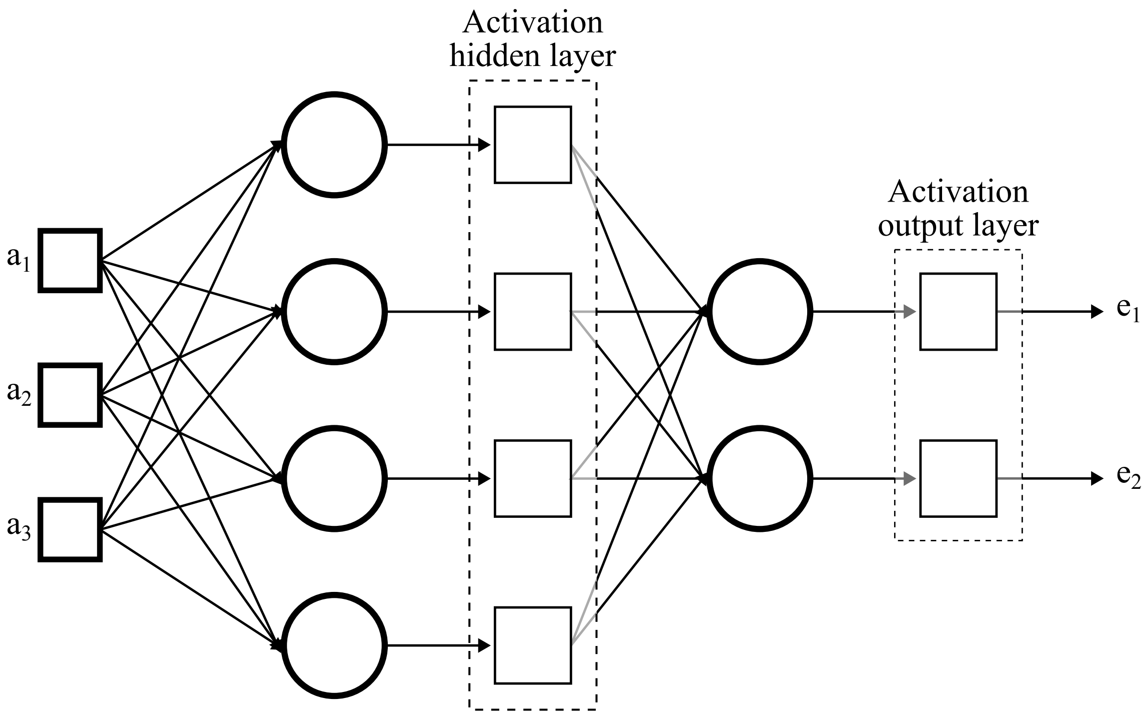
14. Write down the *empirical risk/0-1* loss function for training a classifier (1p). How can a classifier be trained using this loss function (1p)?

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: 2023-03-20
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 2

1. Figure 1 shows the schematic of the backbone of a two-layer fully connected network which could, after appropriate modifications, be able to classify a dataset which has two not-linearly separable classes and which feature space is not zero-centred. Each element in the dataset is represented by a vector with length three. The training shall use on-line back propagation, with the quadratic error function $\epsilon = \|\mathbf{e} - \mathbf{g}\|^2$. Here $\mathbf{g}$ is the desired output of the network and $\mathbf{e}$ the network output.



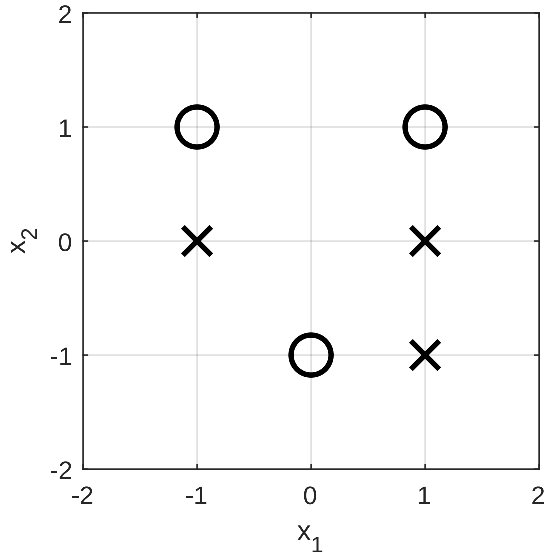
Figur 1: Schematic of a two layer fully connected network.

a) Modify the network design and specify all its components considering the specifications of the dataset (two not-linearly separable classes and which feature space is not zero-centred). Motivate your choices using words. (1p)

b) For the updated version of the network using the choices you made, specify all the variables including their dimensions that you will use to derive the update rule for the weights of the network, and provide the mathematical expression describing the inputs, outputs and the operations in each layer. (1p)

c) Derive the element-wise (not matrix form!) update rules for all the weights in the network. (3p)

2. You have the following data:



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

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

We are using AdaBoost on this data by brute force optimization of 'decision stumps' as weak classifiers. The first iteration yields an optimal weak classifier using feature 2, threshold $\tau_1 = 0.5$, and polarity 1. Sample 2 is misclassified, giving an error

$$\varepsilon_1 = \frac{1}{6} \qquad \alpha_1 = \frac{1}{2}\ln\left(\frac{1-\varepsilon_1}{\varepsilon_1}\right) = \ln\left(\sqrt{5}\right) \approx 0.80$$

and updated normalized weights

$$\mathbf{d}_1 = \begin{bmatrix} \frac{1}{10} & \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \end{bmatrix}$$

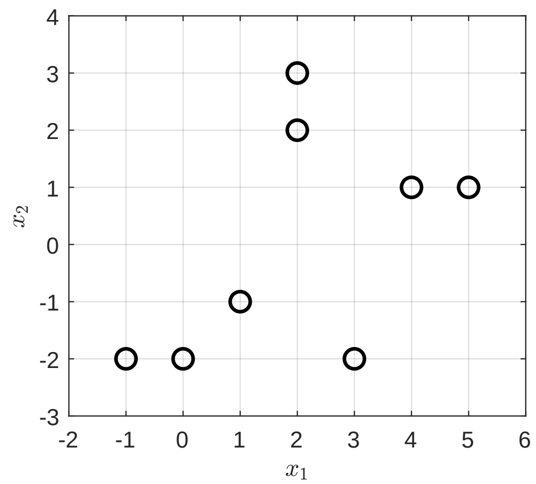
a) Perform the next two iterations of AdaBoost on this data. You do not have to calculate the weight update for the final iteration. (3p)
Hint: Using exact values in the calculations (i.e. no approximations with calculator) will give very clean answers.

b) Draw the decision boundary of the strong classifier defined by the three weak classifiers. (2p)
Hint: This one is probably easier with calculator.

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: 2023-03-20
Course Code: TBMI26 / 732A55	Exam Code: TEN1

3. We have the data



defined by the matrix

$$\mathbf{X} = \begin{bmatrix} -1 & 0 & 1 & 2 & 2 & 3 & 4 & 5 \\ -2 & -2 & -1 & 2 & 3 & -2 & 1 & 1 \end{bmatrix}$$

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:

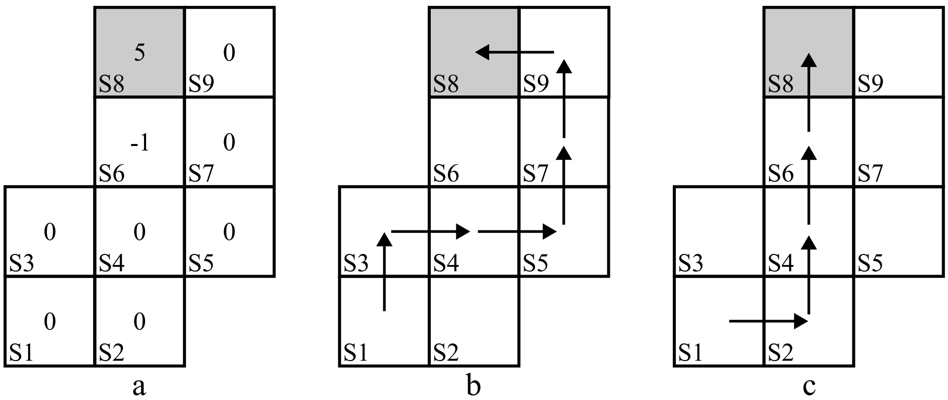
$$\lambda_1 = \frac{T}{2} + \sqrt{\frac{T^2}{4} - D} \text{ and } \lambda_2 = \frac{T}{2} - \sqrt{\frac{T^2}{4} - D},$$

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.

4. Figure 2.a shows a reinforcement learning problem with nine states in which the valid actions are *right*, *left* and *up*. State S_8 is terminal and moving into it results in a reward of 5. Moving into state S_6 results in a reward of -1. Moving into all the remaining states result in a reward of 0.
   Two possible sequences of action are shown in Figure 2.b and 2.c.



Figur 2: All possible states and rewards (a), and sequences of action (b and c).

   Show how the Q-values are modified by the Q-learning algorithm if sequence ***b*** is used once, followed by sequence ***c***, and then a final use of sequence ***b***.

   Give the results as a function of γ and α. All Q-values are initialized at 0. (5p)