Exam in
Neural Networks and Learning Systems
TBMI26 / 732A55

Time:	2022-03-21, 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 2022-04-11 at the latest. The exams will then be available at studerandeexpeditionen at IMT.

GOOD LUCK!

AID:	Exam Date: 2022-03-21
Course Code: TBMI26 / 732A55	Exam Code: TEN1

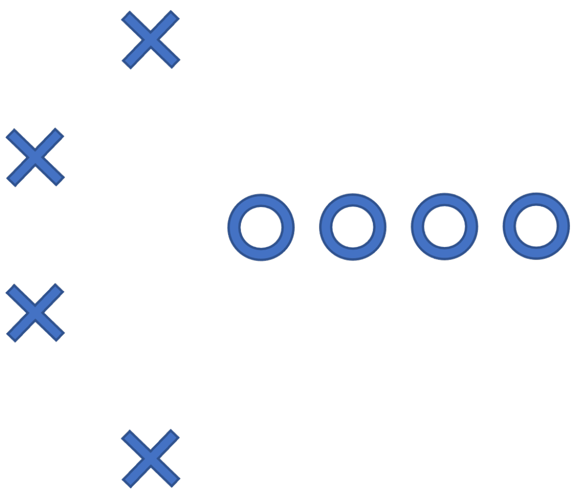
Part 1

One-point questions

1. Which of the following methods are unsupervised learning methods:

   - Mixture of Gaussians
   - Back-propagation
   - k-NN
   - k-means
   - SVN (Support Vector Machines)
   - PCA (Principal Component Analysis)

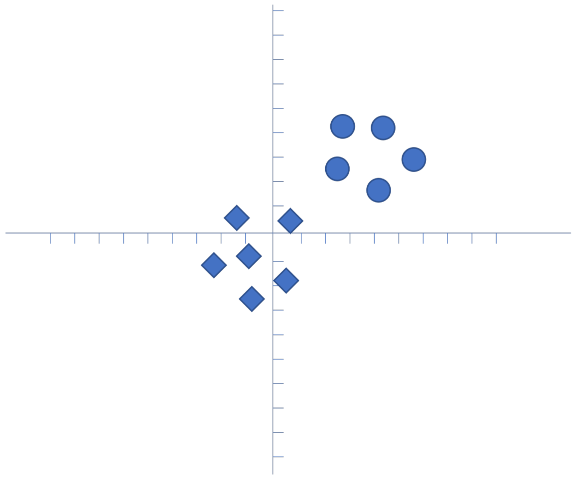
2. Draw the line for which the discriminant function $f(\mathbf{x}) = 0$ for a linear SVN without slack variables and mark the support vectors in the figure below



AID:	Exam Date: 2022-03-21
Course Code: TBMI26 / 732A55	Exam Code: TEN1

3. Assume you are using a single perceptron to solve the classification problem in the figure below, but you have forgotten to add a bias weight.

   Draw the line for which the discriminant function $f(\mathbf{x}) = 0$.



4. What can be said about the correlations between the different principal components of a data distribution?

5. What kind of task is the "U-net" used for?

6. What is the k-means algorithm used for?

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. SVM is essentially a linear classifier. How can this be used to solve non-linearly separable problems?

9. Training a large network for too long can result in overfitting. How can you know when to stop training in order to avoid this?

10. What does the ”value function’ $V(s)$ in reinforcement learning describe?

AID:	Exam Date: 2022-03-21
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Two-point questions

11. What is the ”vanishing gradient” problem, and how can it be avoided?

12. What kind of learning paradigm (supervised, unsupervised or reinforcement learning) does auto-encoders belong to? Mention an application for auto-encoders!

13. Consider a Gaussian kernel function $\kappa(\mathbf{x}_1, \mathbf{x}_2) = e^{-\frac{\|\mathbf{x}_1-\mathbf{x}_2\|^2}{4}}$. What is the distance between two feature vectors

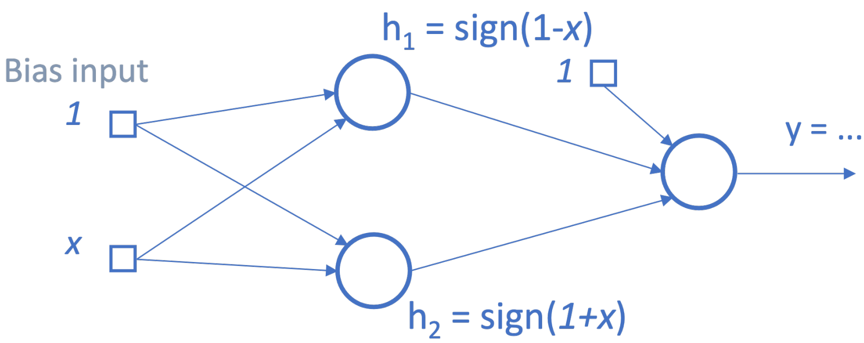
$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \mathbf{x}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

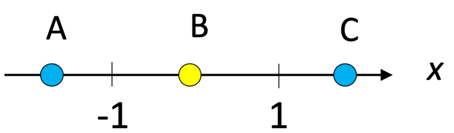
in the new feature space defined by this kernel function?

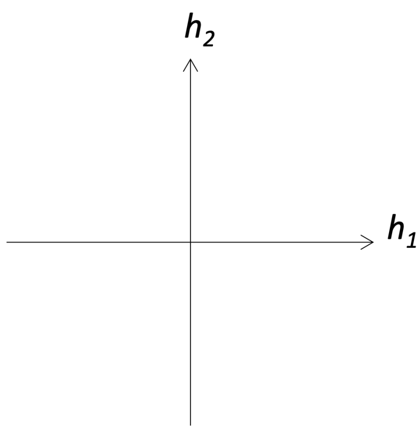
AID:	Exam Date: 2022-03-21
Course Code: TBMI26 / 732A55	Exam Code: TEN1

14. Write a 3-by-3 convolution kernel that detects vertical lines in an image. The kernel should not be sensitive to the mean intensity of the image neighborhood.

15. Consider the network in the figure below. Show where the three datapoints A, B, and C, will be represented the intermediate space $\mathbf{h}$ and define the function $y(h_i)$ in the output layer that separates the two classes. Also draw the separating line in the figure.







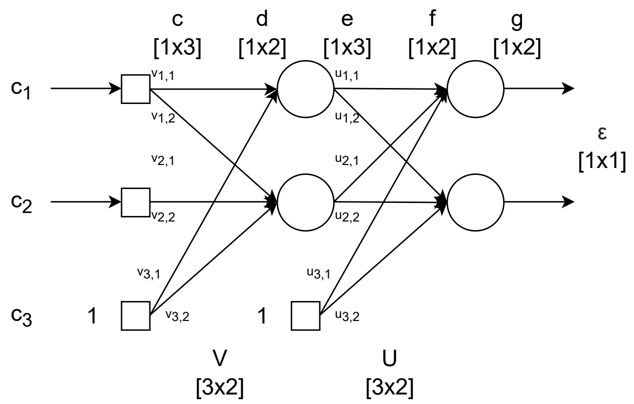
AID:	Exam Date: 2022-03-21
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 2

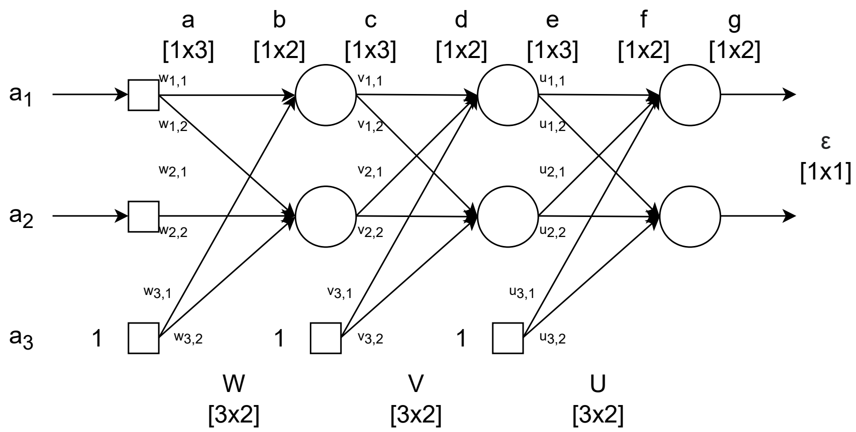
1. a) Consider the neural network in Figure ??. Provide mathematical expressions describing its inputs, outputs and operation of each layer. Use the figure to derive the element-wise weight update rule for online training that would minimize the square error function. (3p)

   b) Do the same for the network in Figure ??. *Tip: You can reuse most of your solution to a).* (2p)

$ReLU(x) = \max(0, x)$



Figur 1: Neural network diagram a).



Figur 2: Neural network diagram b).

AID:	Exam Date: 2022-03-21
Course Code: TBMI26 / 732A55	Exam Code: TEN1

2. You have the following data (see Figure 1):

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

where $\mathbf{X}$ contains eight 2d-samples (one per column), and $\mathbf{y}$ contains classification labels for the corresponding samples. We have performed two iterations of AdaBoost using 'decision stumps' as weak classifiers on the data $\mathbf{X}$. We calculated the following weights $\mathbf{d}$, classification labels $\mathbf{c}$ and α for each weak classifier:

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_3 \end{bmatrix} = \begin{bmatrix} 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 \\ 1/2 & 1/14 & 1/14 & 1/14 & 1/14 & 1/14 & 1/14 & 1/14 \\ 7/22 & 1/22 & 1/22 & 1/22 & 1/22 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

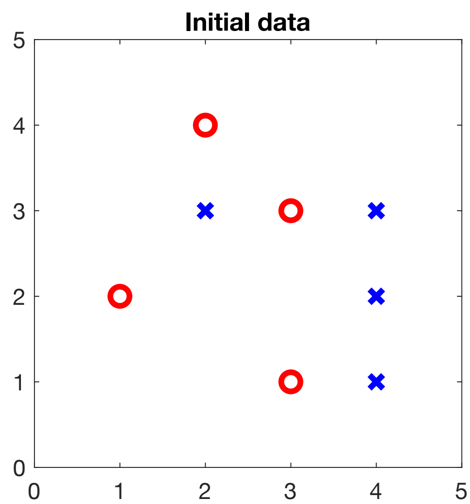
$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 \end{bmatrix}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0.97 \\ 0.65 \end{bmatrix}$$

a) Perform the third AdaBoost iteration on the data $\mathbf{X}$ using the labels $\mathbf{y}$ and 'decision stumps' as weak classifiers. Calculate $\mathbf{d}_4$, $\mathbf{c}_3$ and α_3 (3p)

b) Classify the data $\mathbf{X}$ using strong classifiers based on one, two, and three weak classifiers. Write the accuracy in each case. (2p)

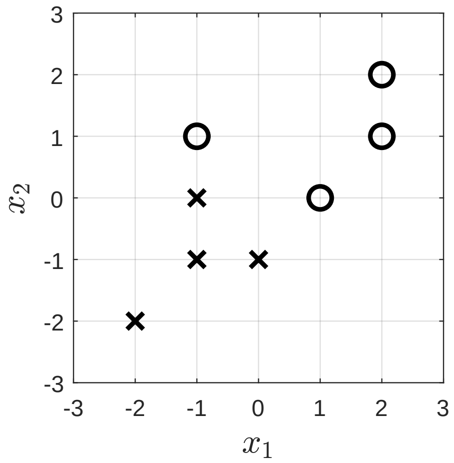
Hint: The standard way of updating the weights in the standard AdaBoost method is $d_{t+1}(i) \propto d_t(i)e^{-\alpha_t y_i h_t(\mathbf{x})}$, where $\alpha_t = \frac{1}{2}\ln\frac{1-\epsilon_t}{\epsilon_t}$.



Figur 3: Blue crosses and red circles represent classes 1 and −1 respectively.

AID:	Exam Date: 2022-03-21
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*:

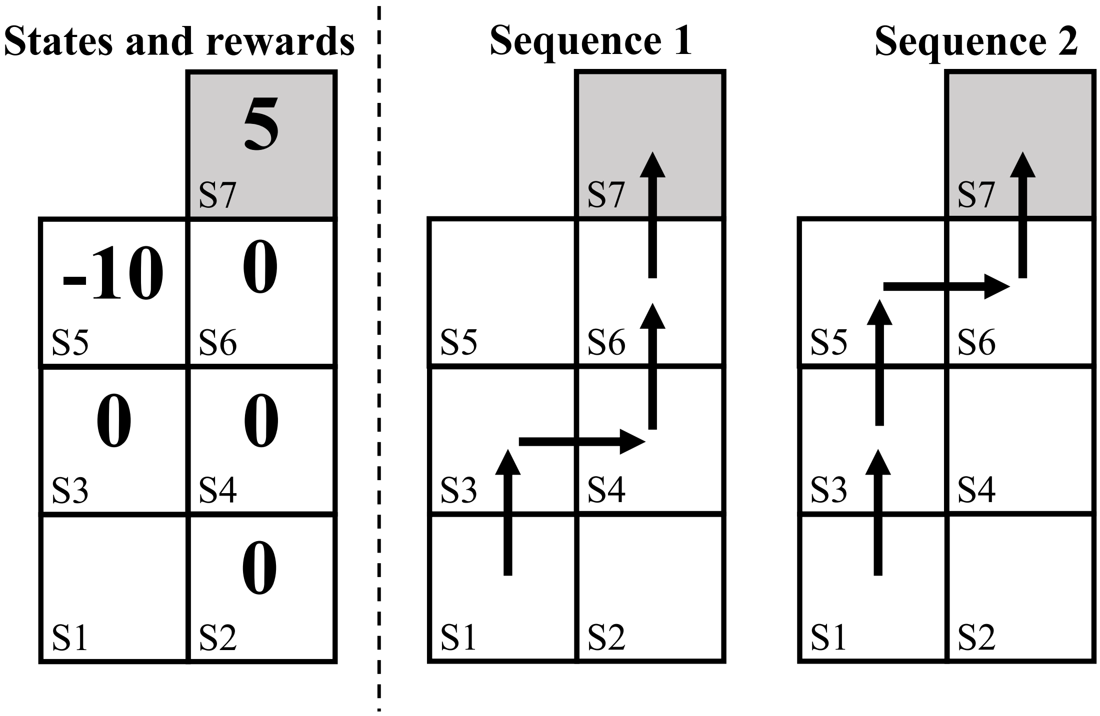


Perform Linear Discriminant Analysis (LDA) on the data to reduce the dimensionality to one dimension that separates the two classes optimally. Draw the reduced data. (5p)

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.

4. Figure 2 shows a reinforcement learning problem with seven states in which the valid actions are *up* and *right*. State S_7 is the final state and moving into it results in a reward of 5. Moving into state S_5 results in a reward of -10. All remaining states result in a reward of 0.
   Two possible sequences of actions (***Sequence 1*** and ***Sequence 2***) are shown in Figure 2.



Figur 4: State model showing all possible states and rewards.

Show how the Q-values of the states are modified by the Q-learning algorithm if ***Sequence 1*** is used once (2p), followed by ***Sequence 2*** two times (3p). Give the results as a function of learning rate $\alpha \in [0, 1]$ and discount factor $\gamma \in [0, 1]$. Moreover, note that all Q-values are initialized with 0 and that $\max_a \hat{Q}(7) = 0$