Exam in
Neural Networks and Learning Systems
TBMI26 / 732A55

Time:	2022-06-09, 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 2022-07-01 at the latest. The exams will then be available at studerandeexpeditionen at IMT.

GOOD LUCK!

AID:	Exam Date: 2022-06-09
Course Code: TBMI26 / 732A55	Exam Code: TEN1

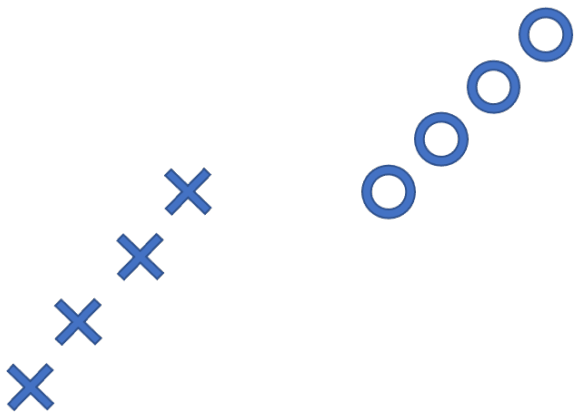
Part 1

One-point questions

1. Machine learning is often divided into the categories Supervised, Unsupervised and Reinforcement learning. Categorize the following three learning methods accordingly:

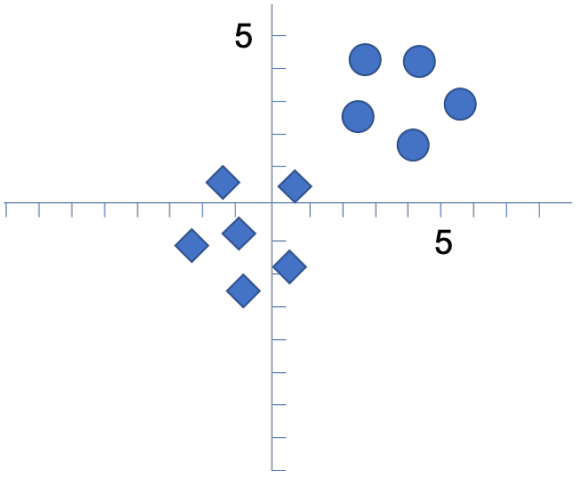
   - Q-leaerning
   - k-means
   - kNN (k nearest neighbors)

2. Draw (approximately) $|f(\mathbf{x})| = 0.5$ where $f(\mathbf{x})$ is the discriminant function in a linear SVN without slack variables.

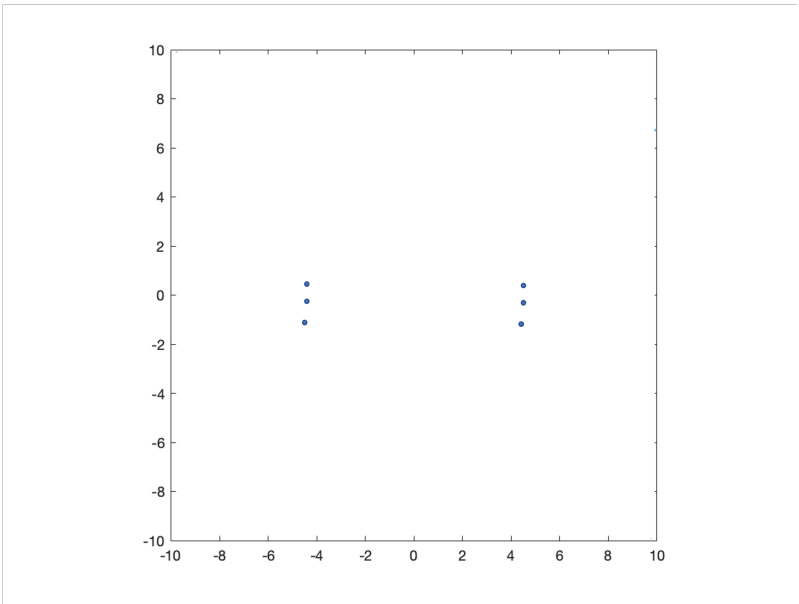


AID:	Exam Date: 2022-06-09
Course Code: TBMI26 / 732A55	Exam Code: TEN1

3. Assume you are using a single perceptron to solve the classification problem in the figure below. Write the weights **w** that separates the classes. (The weights do not have to be exact.)

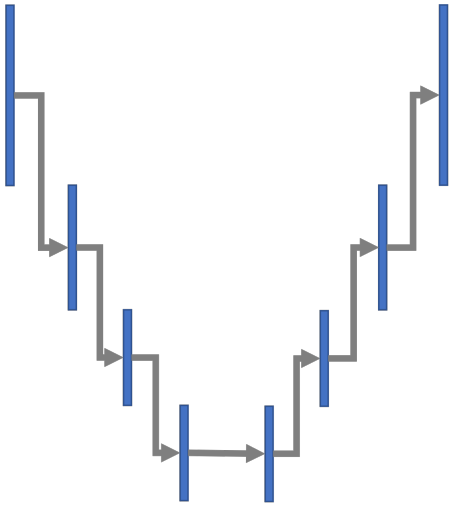


4. Draw the first principal component of the distribution in the figure below!



AID:	Exam Date: 2022-06-09
Course Code: TBMI26 / 732A55	Exam Code: TEN1

5. What is missing in this illustration of the "U-net"?



6. Assume you have a set of data points in a two-dimensional feature space and you want to cluster the points into three clusters using Mixture of Gaussians. How many parameters (scalar numbers) need to be estimated in the optimization, disregarding the set membership variables S_i?

7. What kind of operation is often used between the convolutional layers in a CNN in order to create shift invariance and reduce the size of the feature maps?

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

AID:	Exam Date: 2022-06-09
Course Code: TBMI26 / 732A55	Exam Code: TEN1

9. Why isn't it optimal to try to reach the lowest possible error on the training data when traing a neural network?

10. 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: 2022-06-09
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Two-point questions

11. Consider the product

$$\prod_{k=1}^{K} \sigma'(s_k).$$

What happens when K is large and $0 \leq \sigma'(s) \leq 1$ and why is this relevant in machine learning? Suggest a σ that alleviates this effect.

12. Describe the difference between a deep neural network used for image segmentation and one that is used for image classification in terms of the network architecture.

13. Consider the kernel function

$$\kappa(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{1 + ||\mathbf{x}_1 - \mathbf{x}_2||^2}$$

What is the distance between two feature vectors

$$\mathbf{x}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \text{ and } \mathbf{x}_2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

in the new feature space defined by this kernel function?

AID:	Exam Date: 2022-06-09
Course Code: TBMI26 / 732A55	Exam Code: TEN1

14. Write a 3-by-3 convolution kernel that detects small dots (i.e. bright pixels on dark background or vice versa) in an image. The kernel should not be sensitive to the mean intensity of the image neighborhood.

15. Draw a decision tree that implements the following discriminant function:

$$y = \begin{cases} -1, \text{ for } (x < -1) \text{ OR } (x > 1) \\ 1, \text{else} \end{cases}$$

AID:	Exam Date: 2022-06-09
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 2

1. You have the following training samples:

$$\begin{aligned} \mathbf{x} &= [\,\text{-1} \quad \text{-0.9} \quad \text{-0.3} \quad \text{-0.1} \quad 0.1 \quad 0.3 \quad 0.9 \quad 1.0 \quad 1.2 \quad 1.5\,]^T \\ \mathbf{d} &= [\,\text{-1} \quad \text{-1} \quad \text{-1} \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad \text{-1} \quad \text{-1}\,]^T \end{aligned}$$

   Where $\mathbf{d}$ is the desired output for the samples in $\mathbf{x}$.

   a) Your task is to use a neural network setup in order to separate the data using Cover's theorem. You shall have one node in the output layer. Use *sign* as activation function, both in the hidden layer and the output layer. Sketch the network and assign $\mathbf{W}$ (weights in the hidden layer) and $\mathbf{V}$ (weights in the output layer) so that the output gives 100% in accuracy (4p).

   b) What would you have to initialize and/or change in order for a computer to solve the task above using backpropagation with gradient descent (1p)?

AID:	Exam Date: 2022-06-09
Course Code: TBMI26 / 732A55	Exam Code: TEN1

You have the following data:

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

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

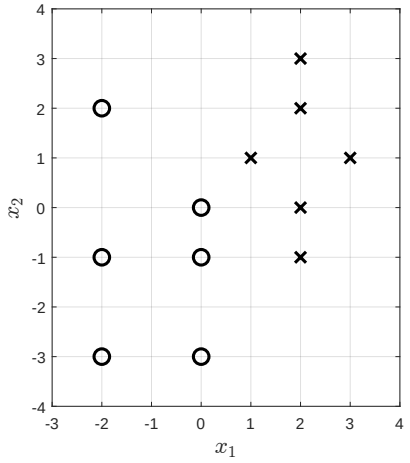
2. a) Perform the first AdaBoost iteration on the data $\mathbf{X}$ using the labels $\mathbf{y}$ using brute force optimization and 'decision stumps' as weak classifiers. (4p)

   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: 2022-06-09
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*:



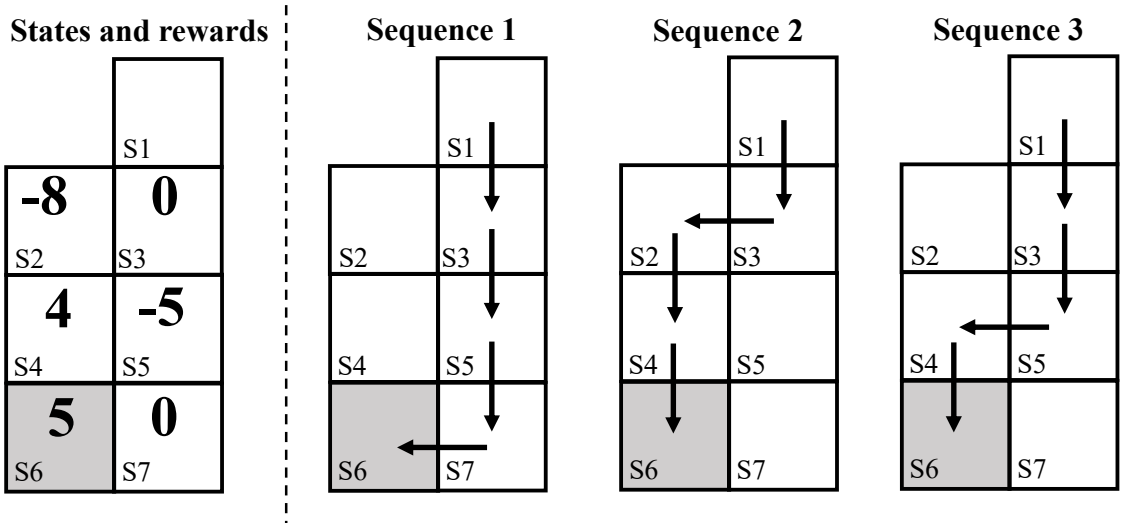
Perform Linear Discriminant Analysis (LDA) on the data to reduce the dimensionality to one dimension that separates the two classes optimally. Draw the reduced data. (5p)

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.

AID:	Exam Date: 2022-06-09
Course Code: TBMI26 / 732A55	Exam Code: TEN1

4. Figure 1 shows a reinforcement learning problem with seven states in which the valid actions are *down* and *left*. State S_7 is the final state and moving into it results in a reward of 5. Moving into state S_2 results in a reward of -10, moving into state S_5 results in a of -5, moving into S_4 in a positive reward of 4 and reaching the final state S_6 in a reward of 5. All remaining states result in a reward of 0.
   Three possible sequences of actions (***Sequence 1***, ***Sequence 2*** and ***Sequence 3***) are shown in Figure 1.



Figur 1: 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 (1p), followed by ***Sequence 2*** (2p) and then by ***Sequence 3*** (2p). 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 the value of S_6 is 0.