

Exam in Neural Networks and Learning Systems TBMI26 / 732A55

Time: 2019-06-12, 14-18
Teacher: Magnus Borga, Phone: 013-286777
Allowed additional material: Calculator, English dictionary

Read the instructions before answering the questions!

The exam consists of three parts:

- Part 1 Consists of ten questions. The questions test general knowledge and understanding of central concepts in the course. The answers should be short and given on the blank space after each question. Any calculations do **not** have to be presented. Maximum one point per question.
- Part 2 Consists of five questions. These questions can require a more detailed knowledge. Also here, the answers should be short and given on the blank space after each question. Only requested calculations have to be presented. Maximum two points per question.
- Part 3 Consists of four questions. All assumptions and calculations made should be presented. Reasonable simplifications may be done in the calculations. **All calculations and answers on part 3 should be on separate papers! Do not answer more than one question on each paper!** Each question gives maximum five points.

The maximum sum of points is 40 and to pass the exam (grade 3) normally 18 points are required. There is no requirement of a certain number of points in the different parts of the exam. The answers may be given in English or Swedish. **Write clearly using block letters! (Do not use cursive writing.) Answers that are difficult to read, will be dismissed.**

The result will be reported at 2019-06-28 at the latest. The exams will then be available at "studerandeexpeditionen" at IMT.

GOOD LUCK!

AID:	Exam Date: 2019-06-12
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 1

(Please use the space under each question for your answer in this part.)

1. Classify the following learning methods as supervised (S) or unsupervised (U):

- k-Nearest Neighbors
- Support Vector Machines
- AdaBoost
- Principal Component Analysis
- Multi-layer Perceptron (Neural Network)
- Mixture of Gaussian Clustering

2. How is the accuracy of a classifier calculated?

3. Why are the following two functions not useful in the hidden layers in a back-propagation neural network?

- $y = s$
- $y = \text{sign}(s)$

AID:	Exam Date: 2019-06-12
Course Code: TBMI26 / 732A55	Exam Code: TEN1

8. In which kind of learning tasks is linear units more useful than sigmoid activation functions in the output layer of a multi-layer neural network?
9. Explain the purpose with the so-called *slack variables* in Support Vector Machines.
10. All the weights one layer in a neural network can be described as a matrix W . Describe an important property of this matrix for a convolutional layer in a CNN.

AID:	Exam Date: 2019-06-12
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 2

(Please use the space under each question for your answer in this part.)

11. What are the two (main) differences between supervised and reinforcement learning?

12. Describe briefly the two steps that are iterated in the k-means clustering algorithm.

13. Consider the following explicit non-linear mapping of the input data $\mathbf{x} = (x_1, x_2)^T$:

$$\boldsymbol{\varphi}(\mathbf{x}) = \begin{pmatrix} \varphi_1(\mathbf{x}) \\ \varphi_2(\mathbf{x}) \\ \varphi_3(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2} \cdot x_1 x_2 \end{pmatrix} \quad (1)$$

Kernel methods use kernel functions to avoid explicit mappings and calculations in higher-dimensional feature spaces as above.

- What does a kernel function $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ calculate? (1p)
- Find the expression for the kernel function that corresponds to the mapping above. (1p)

AID:	Exam Date: 2019-06-12
Course Code: TBMI26 / 732A55	Exam Code: TEN1

14. ReLU activation functions are more and more used in neural networks instead of the tanh activation function. Draw both activation functions and give a) an advantage of the ReLU function compared to the tanh function. b) a disadvantage of the ReLU function compared to the tanh function.

15. In SVM, a cost function is minimized under the following constraint:

$$d_i(\mathbf{w}^T \mathbf{x}_i) + b \geq 1.$$

What is the cost function being minimized, and for which \mathbf{x}_i are the constraint fulfilled with equality?

AID:	Exam Date: 2019-06-12
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 3

(N.B. Write all answers in this part on separate sheets of papers! Don't answer more than one question on each sheet!)

16. The convolution of a 2D image $f(x, y)$ and a kernel $h(x, y)$ is defined as

$$g(x, y) = (f * h)(x, y) = \sum_{\alpha=-\infty}^{\infty} \sum_{\beta=-\infty}^{\infty} f(\alpha, \beta) h(x - \alpha, y - \beta).$$

- a) Perform the convolution below, i.e. calculate the image C. All values outside the image array A are equal to zero. In the arrays A and B, the respective number written in bold face is at position $(x, y) = (0, 0)$. Note that C is only a part of the convolution result. (2p)

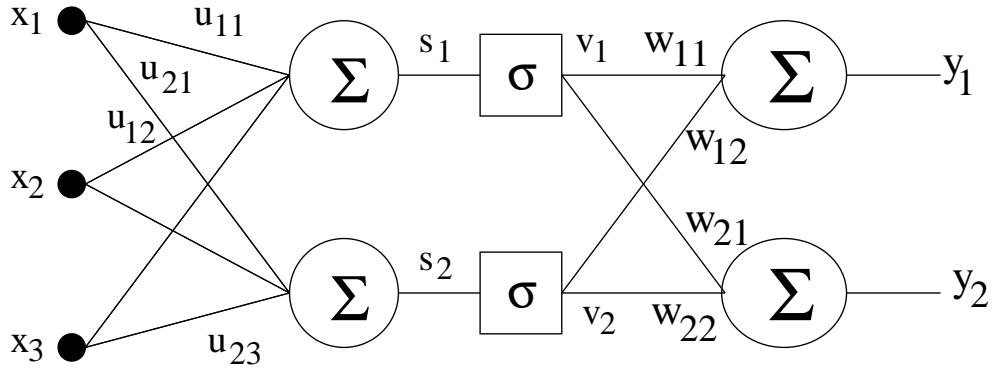
0	1	0	2	*	0	1	0	=	?	?	?	?
1	0	1	0		1	1	2		?	?	?	?
0	2	0	0		0	2	0		?	?	?	?

image A kernel B = image C

- b) In practice, when implementing convolution, e.g. like `convolve(A,B)`, no part of the kernel can be placed outside the image. Consequently, the resulting 2D array has size 1×2 and is equal to the central part of the image array C above. By extending the image array A to a new image array AA in a suitable way, `convolve(AA,B)` will be equal to C. Give the image array AA. (1p)
- c) A CNN consists of N complex layers. Each complex layer consists of a convolution with a 3×3 kernel, a sigmoid activation, and a max pooling layer with stride 2 in each dimension. The input image is 512×512 . Compute the (spatial) size of the layer after the first and second pooling. When is the image of size 1×1 , i.e. what is the maximum value for N ? (2p)

AID:	Exam Date: 2019-06-12
Course Code: TBMI26 / 732A55	Exam Code: TEN1

17. We want to classify some data using the network below. There are three input parameters and two output classes. The network is trained using standard error back-propagation, i.e. the square error should be minimized using gradient search. The network has one hidden layer and one output layer, see figure.



A team of engineers implemented the above neural network, but forgot the bias weight and the activation function. So, without knowing it, they used $\sigma(x) = x$ as activation function.

- Derive the update expressions for the weights in both layers. (3p)
- Explain the restrictions, if any, that are imposed upon the decision boundaries because of the nonexistent bias weights and activation functions. (2p)

AID:	Exam Date: 2019-06-12
Course Code: TBMI26 / 732A55	Exam Code: TEN1

18. You have the following data:

$$\mathbf{X}_{Train} = \begin{bmatrix} 1 & 2 & 2 & 2 & 3 & 4 \\ 1 & 1 & 2 & 3 & 3 & 3 \end{bmatrix} \quad \mathbf{y}_{Train} = [1 \quad -1 \quad -1 \quad 1 \quad 1 \quad -1]$$

$$\mathbf{X}_{Test} = \begin{bmatrix} 0 & 1 & 2 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 & 3 & 3 \end{bmatrix} \quad \mathbf{y}_{Test} = [1 \quad -1 \quad -1 \quad 1 \quad 1 \quad -1]$$

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

$$\mathbf{d} = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/4 & 1/8 & 1/8 & 1/8 & 1/8 & 1/4 \\ 1/6 & 1/12 & 1/12 & 1/4 & 1/4 & 1/6 \end{bmatrix}$$

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

$$\alpha = \begin{bmatrix} 0.35 \\ 0.55 \end{bmatrix}$$

- Perform the third AdaBoost iteration on the data \mathbf{X}_{Train} using the labels \mathbf{y} and 'decision stumps' as weak classifiers. Calculate \mathbf{d}_4 , \mathbf{c}_3 and α_3 (3p)
- Apply the strong classifier on the data \mathbf{X}_{Train} and the data \mathbf{X}_{Test} . What is the accuracy for the \mathbf{X}_{Train} data and the \mathbf{X}_{Test} respectively? (1p)
- Will we be able to achieve an accuracy of 100 % for the test data with more iterations? Motivate your answer! (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: 2019-06-12
Course Code: TBMI26 / 732A55	Exam Code: TEN1

19. Figure 1 shows a reinforcement learning problem with eight states in which the valid actions are *right* 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. All remaining states result in a reward of 0.

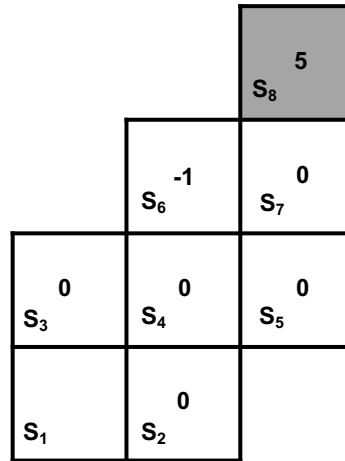


Figure 1: All possible states and rewards

Two possible sequences of action are shown in Figure 2.

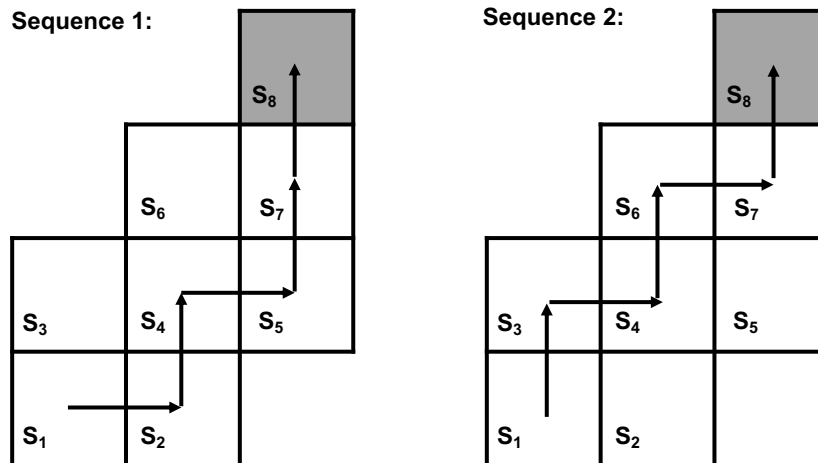


Figure 2: Sequences of action

Show how the Q-values are modified by the Q-learning algorithm if sequence 1 is used once, followed by sequence 2, and then a final use of sequence 1.

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