Exam in
Neural Networks and Learning Systems
TBMI26 / 732A55

Time:	2019-08-31, 08-12
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 2019-09-17 at the latest. The exams will then be available at "studerandeexpeditionen" at IMT.

GOOD LUCK!

AID:	Exam Date: 2019-08-31
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 1

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

1. Machine learning can be divided into supervised, unsupervised and reinforcement learning. Mention one example of learning methods for each of these three classes.

2. What happens if the input to the bias weight in a perceptron is set to the constant value 2?

3. Connect the concepts:
   1. Maximum margin principle
   2. Back propagation
   3. Brute force optimization

   A. Neural Networks
   B. Decision Tree
   C. Support Vector Machine

4. What is the difference between k-means clustering and Mixture of Gaussians clustering?

5. In Linear Discriminant Analysis, a quotient between a distance d and a variance v is maximized. What is described by d and v?

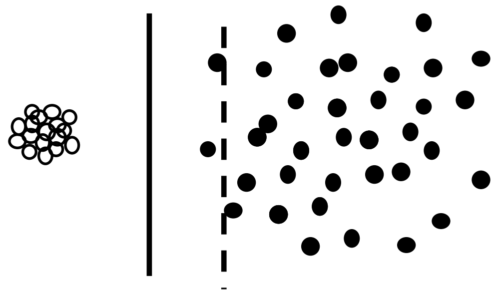
6. Explain the principle of ϵ-*greedy exploration* in the context of reinforcement learning.

7. The following update rule is sometimes used in gradient search:

$$\Delta w_{ji}(t) = \alpha \Delta w_{ji}(t-1) - \frac{\delta \epsilon(t)}{\delta w_{ji}}$$

What is the term $\alpha \Delta w_{ji}(t-1)$ called?

8. In the figure below, a simple one-layer perceptron has been used for classification. Give an explanation in terms of the activation function for the different resulting discriminant functions (solid vs dashed line).



9. What output will you get if you use one of the support vectors as input to a Support Vector Machine?

10. Mention a type of neural networks that uses weight sharing (parameter sharing).

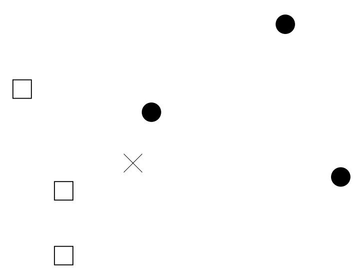
AID:	Exam Date: 2019-08-31
Course Code: TBMI26 / 732A55	Exam Code: TEN1

Part 2

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

11. What does the "credit assignment problem" refer to in the context of reinforcement learning?

12. Which class does X belong to using a kNN-classifier using $k = 1$ and $k = 3$? why? How can we handle $k = 2$?



13. Consider a polynomial kernel function $\kappa(\mathbf{x}_1, \mathbf{x}_2) = 1 + (\mathbf{x}_1^T \mathbf{x}_2)^2$. What is the distance between two feature vectors

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

in the new feature space defined by this kernel function?

14. When training a supervised learning system, the available data are usually divided in three groups that are used for different purposes. Name these groups and explain how they are used!

15. Explain what a decision stump is and how you can train it.

AID:	Exam Date: 2019-08-31
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 that one question on each sheet!)

16. You have the following training samples:

X	=	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
D	=	-1	-1	1	1	1	1	-1	-1	-1

where D is the desired output for the samples in 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 W (weights in the hidden layer) and 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)?

17. PCA can be formulated as an optimization problem. Formulate a suitable cost function (1p) and show that its optimum corresponds to the first eigenvector of the data covariance matrix (4p).

    (Assume a zero mean so that the covariance matrix can be written like $\mathbf{C} = E[\mathbf{x}\mathbf{x}^T]$.)

AID:	Exam Date: 2019-08-31
Course Code: TBMI26 / 732A55	Exam Code: TEN1

18. Assume the following six 2d-samples (written as columns): $\mathbf{X} = \begin{bmatrix} -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 \end{bmatrix}$.

    Assume we have two prototypes (k=2): $\mathbf{p}_1 = \begin{bmatrix} -0.5 \\ 0 \end{bmatrix}$ and $\mathbf{p}_2 = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$.

    a) Perform two iterations of k-Means. Explain and sketch the solution steps and the reasoning behind them. (2p)

    b) Perform one iteration of Mixture of Gaussians (MoG). Assume initial covariance matrices $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{I}$ for the prototypes. Explain and sketch the solution steps and the reasoning behind them. The exact numerical solution is not important, i.e. you don't have to calculate the covariance matrices or the likelihoods explicitly. However, the general shape of the updated Gaussians should be drawn.(3p)

    N-dim. normal distribution:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \mathbf{C}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right]$$

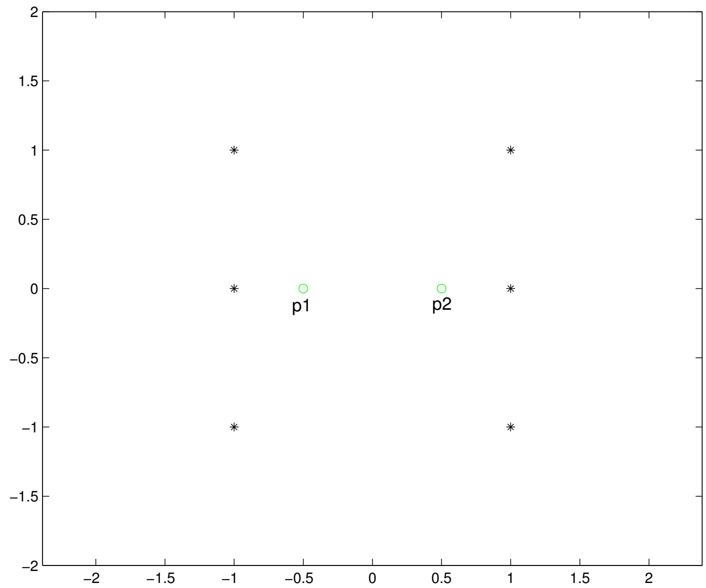


Figure 1: The six samples (stars) and the initial values of the two prototype vectors (circles).

19. The figure shows two different deterministic state models and the corresponding rewards. The states are enumerated and the arrows represent actions. The numbers close to the arrows show the corresponding rewards. If the system reaches a state denoted "End" no additional rewards are given, i.e. the V-function is defined as 0 in such a state. An optimal policy is in this context the policy which maximizes the reward.

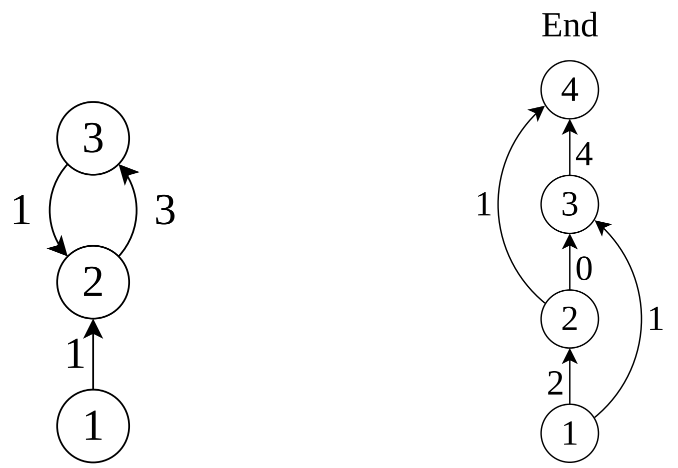


Figure 2: The state models A and B.

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

b) Calculate the optimal Q- and V-functions for system B as functions of $0 < \gamma < 1$. (3p)