Exam in
Neural Networks and Learning Systems
- TBMI26

Time:	2014-03-20, 14-18
Teacher:	Ola Friman, Phone: 0701-475553
Allowed additional material:	Calculator, English dictionary

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 should be on separate papers. 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.

The result will be reported at 2014-04-03 at the latest. The exams will then be available at IMT.

GOOD LUCK!

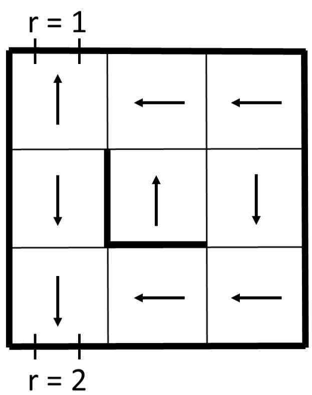
AID:	Exam Date: 2014-03-20
Course Code: TBMI26	Exam Code: TEN1

Part 1

1. An algorithm that detects faces is common in modern cameras and mobile phones. Is such a function typically trained using *supervised learning*, *unsupervised learning* or *reinforcement learning*?

2. Mention a reason why correlated features are undesired in supervised learning.

3. A learning system is exploring a room to find the best way to either of the two exits according to the drawing below. Note that the exits give different rewards. In the context of reinforcement learning, give the value function for each position in the room for the movement policy given by the arrows. Use the discount factor 0.5.



4. Which trick is used in *deep learning* of a neural network to train many layers?

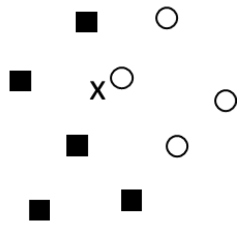
5. Describe the kind of optimization problems for which Genetic algorithms may be a suitable method.

6. Describe or draw a situation when the so-called *slack* variables are required in Support Vector Machines.

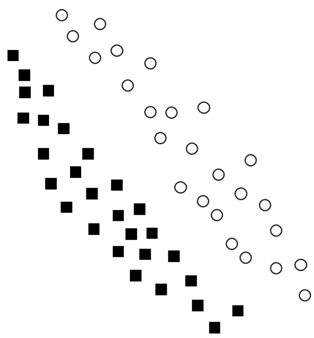
7. After you have trained an AdaBoost classifier you note that one training sample has received a larger weight than other training samples. What could be the reason and effect of this?

8. Kernel methods have a main disadvantage when we have lots of training data. Which?

9. Which class does the data sample 'x' belong to using a k-Nearest Neighbor classifier with $k = 1$ and $k = 3$ respectively?



10. You want to perform a dimensionality reduction from 2 dimensions to 1 dimension of the 2-class data below. Draw the approximate the projection directions you would obtain using Principal Component Analysis and Fisher Linear Discriminant.



AID:	Exam Date: 2014-03-20
Course Code: TBMI26	Exam Code: TEN1

Part 2

11. Assume that the skill of a learning system is described by the function $\epsilon(w) = 5 - (w - 3)^2$, where w is a parameter we want to optimize so as to get a good performance of the system. In the course, we have seen three optimization approaches:

    - Algebraic optimization
    - Gradient ascent/descent optimization
    - Brute force optimization

    Illustrate how we can find a good value for w using each of these methods!

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

13. Draw a decision tree that solves the XOR problem.

14. A so-called *quant* works in the financial business, applying mathematical models and data analysis methods like the ones in this course to, for example, predict stock markets for trading. Assume that you want to classify if a stock's price next week will be higher or lower than today using a neural network. You have access to historical data of stock prices of the same and similar companies, as illustrated in the figure below.

    - What would be your neural network design? Activation functions? Number of hidden and output nodes?
    - What would you use as features and input to the network? Pre-processing?
    - How would you prevent overfitting?

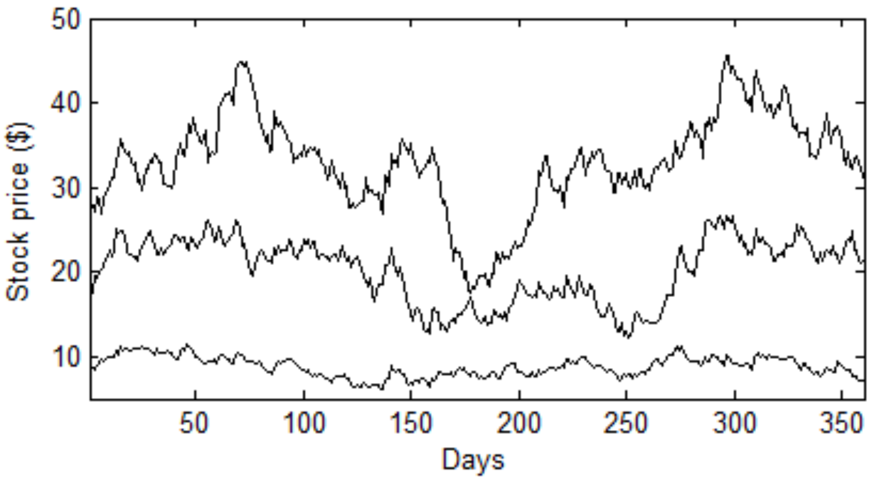
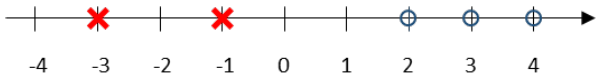


Figure 1: Historical stock prices of different companies. The figure is only for illustration, you are not required to use data from it.

AID:	Exam Date: 2014-03-20
Course Code: TBMI26	Exam Code: TEN1

15. Consider the 1-dimensional problem illustrated in the figure below, with two classes indicated by circles (o) and crosses (x). Draw the classification function we would obtain using *linear SVM* (1p) and motivate the answer using the cost function optimized in linear SVM (1p).



Part 3

16. We have a robot which lives in a world with 4×4 states. Our goal is to train the robot to find its way home in this world *using the shortest path.*

    The states are coded using two indices and are arranged like the table below:

$\mathbf{s} = (1,1)^T$	$\mathbf{s} = (1,2)^T$	$\mathbf{s} = (1,3)^T$	$\mathbf{s} = (1,4)^T$
$\mathbf{s} = (2,1)^T$	$\mathbf{s} = (2,2)^T$	$\mathbf{s} = (2,3)^T$	$\mathbf{s} = (2,4)^T$
$\mathbf{s} = (3,1)^T$	$\mathbf{s} = (3,2)^T$	$\mathbf{s} = (3,3)^T$	$\mathbf{s} = (3,4)^T$
$\mathbf{s} = (4,1)^T$	$\mathbf{s} = (4,2)^T$	$\mathbf{s} = (4,3)^T$	$\mathbf{s} = (4,4)^T$

    Your goal is to design a Q-learning problem which will give the robot the behaviour we desire. The home of the robot is in state $(1,4)$ and the robot will start at a random location each iteration. For the sake of simplicity we only allow the robot to move either right, down or up, so the left action is forbidden.

    Questions:

    a) Design a reward function $r(\mathbf{s}, a)$ which enforces the desired behaviour, i.e. the reward of being in state $\mathbf{s}$ and taking action a. (1p)

    b) Given the reward above, calculate the expected Q-function $Q^*(s_1, s_2, a)$. You are free to set V(1,4) but the discount factor is set to 1. Answer in the format below: (3p)

    Right:

$Q^*(1,1,\text{right})$	$Q^*(1,2,\text{right})$	$Q^*(1,3,\text{right})$	
$Q^*(2,1,\text{right})$	$Q^*(2,2,\text{right})$	$Q^*(2,3,\text{right})$	
$Q^*(3,1,\text{right})$	$Q^*(3,2,\text{right})$	$Q^*(3,3,\text{right})$	
$Q^*(4,1,\text{right})$	$Q^*(4,2,\text{right})$	$Q^*(4,3,\text{right})$	

    Up:

$Q^*(2,1,\text{up})$	$Q^*(2,2,\text{up})$	$Q^*(2,3,\text{up})$	$Q^*(2,4,\text{right})$
$Q^*(3,1,\text{up})$	$Q^*(3,2,\text{up})$	$Q^*(3,3,\text{up})$	$Q^*(3,4,\text{right})$
$Q^*(4,1,\text{up})$	$Q^*(4,2,\text{up})$	$Q^*(4,3,\text{up})$	$Q^*(4,4,\text{up})$

    Down:

$Q^*(1,1,\text{down})$	$Q^*(1,2,\text{down})$	$Q^*(1,3,\text{down})$	
$Q^*(2,1,\text{down})$	$Q^*(2,2,\text{down})$	$Q^*(2,3,\text{down})$	$Q^*(2,4,\text{down})$
$Q^*(3,1,\text{down})$	$Q^*(3,2,\text{down})$	$Q^*(3,3,\text{down})$	$Q^*(3,4,\text{down})$

    c) Calculate the V-function, given your reward and that the discount factor is set to 1. Answer as below: (1p)

$V^*(1,1)$	$V^*(1,2)$	$V^*(1,3)$	$V^*(1,4)$
$V^*(2,1)$	$V^*(2,2)$	$V^*(2,3)$	$V^*(2,4)$
$V^*(3,1)$	$V^*(3,2)$	$V^*(3,3)$	$V^*(3,4)$
$V^*(4,1)$	$V^*(4,2)$	$V^*(4,3)$	$V^*(4,4)$

    Hint:

    $$\begin{aligned} Q^*(x,a) &= r(x,a) + \gamma V^*(g(x, \mu^*(x))) \\ &= r(x,a) + \gamma \max_b Q^*(g(x, \mu^*(x)), b) \end{aligned}$$

AID:	Exam Date: 2014-03-20
Course Code: TBMI26	Exam Code: TEN1

17. To classify a two-dimensional signal we want to use a two layer neural network (one hidden and one output layer) according to figure 2, i.e., two hidden neurons and two output neurons.

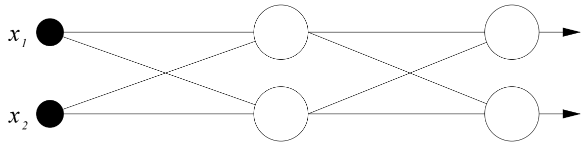


Figure 2: The network for this assignment.

We have two classes coded as $y_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $y_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ respectively. We want to use the activation function $\sigma(x) = \tanh(x)$ for *both* layers. We also want to train the network on-line using gradient descent and with the standard quadratic error:

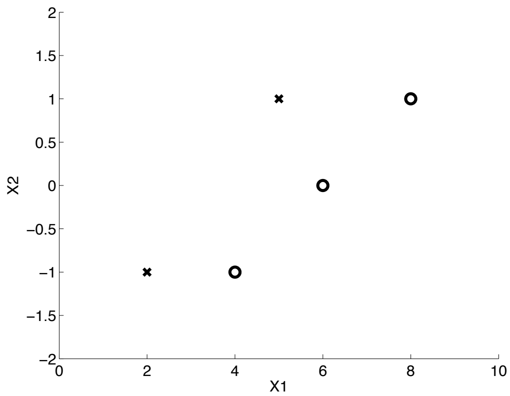
$$\mathcal{E} = \sum_i (y_i - u_i)^2.$$

a) Draw a complete network where you define all signals and weights that are missing in the figure and/or you want to use for your calculations below. (1p)

b) Derive the update rule for all weights in the complete network. Observe that both layers have an activation functions. (4p)

AID:	Exam Date: 2014-03-20
Course Code: TBMI26	Exam Code: TEN1

18. The data points below have two features (x_1 and x_2) and belong to either the class "crosses" or the class "spheres":



a) We think the data volume is too large. Show how the data volume can be reduced to half the size by using principal component analysis. Also, plot the resulting data volume after the reduction. (4p)

b) Is the resulting dimensionality-reduced data good for classification? Motivate your answer! (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

$$\mathbf{e}_1 = \begin{pmatrix} \lambda_1 - d \\ c \end{pmatrix} \text{ and } \mathbf{e}_2 = \begin{pmatrix} \lambda_2 - d \\ c \end{pmatrix}$$

19. The Sneaky Brothers have broken into the mansion of one of their wealthy neighbors. They intended to steal seven precious items, but the things are heavier than they thought and they must leave some stuff behind. The brothers want to maximize the value of the things they take with them. They decide to use a genetic algorithm in order to choose which of the items to steal.

    The weight and value of the items are presented in the table below.

Item	Weight (kg)	Value (kr)
1	5	200
2	10	1000
3	5	300
4	3	400
5	1	200
6	6	100
7	3	300

    The brothers can only carry a maximum of 20 kg together without raising the suspicion of the neighborhood.

    Your task is to help the brothers design and use the genetic algorithm.

    a) Describe a suitable solution representation, a 'fitness"-function and how the crossover and mutation operations can be performed. (3p)

    b) Start with a population of three individuals and run the algorithm one iteration so the next generation is created. Assume a reproduction step where the two best fitted individuals are copied unaltered and they together create a new individual through the crossover operation you suggested in the first part of the question. (2p)