Exam in
Neural Networks and Learning Systems
- TBMI26

Time: 2009-03-14 at 8-12
Teacher: Magnus Borga, Phone: 070-8387534
Allowed additional material: Calculator, Tefyma, Beta, Physics handbook

The exam consist 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 does 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 (not in the exam). 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 2009-03-27 on the latest. The exams will then be available at IMT.

GOOD LUCK!
1. Machine learning can be divided into three different classes. Which?

2. How can one avoid that the discriminant function in a perceptron goes through the origin in the input space?

3. What is determined by the parameter $k$ in $k$-NN?

4. Draw the principal components in the figure below!

5. We have 10 samples in a 20-dimensional space that we want to analyse with a kernel method. How large is the kernel matrix?
6. The figure below shows two types of errors as functions of time. What are these errors?

![Graph showing two types of errors as functions of time.](image)

7. Describe (by an equation) how a momentum term is used in stochastic gradient search.

8. Why isn’t it always best to choose the response with the best Q-value in Q-learning?

9. What are the three most common/important genetic operators?

10. What is the elongated part of the neuron that transmits the output signal called?
11. The figure below shows data for an one-dimensional classification problem along the x-axis. Draw the output signal (as a function of x) and decision boundaries for a linear network that solves the classification problem. Input signals to the network are 1, x and $x^2$. Let output $> 0$ indicate circles and output $\leq 0$ indicate crosses.

12. Consider the following non-linear mapping of the input data $x$:

$$
\varphi_1(x) = x_1^2 \\
\varphi_2(x) = x_2^2 \\
\varphi_3(x) = \sqrt{2} \cdot x_1 x_2
$$

You want to analyse this data with a kernel method. How is the scalar product $\varphi(x_1)^T \varphi(x_2)$ expressed in the input data space?
13. Two signals $x(t)$ and $y(t)$ are created by one common (one-dimensional) signal $s(t)$ and two (one-dimensional) noise components $n_1(t)$ and $n_2(t)$ that are mutually uncorrelated:

$$x(t) = s(t)[-1, -1]^T + n_1(t)[-1, 1]^T \quad \text{and} \quad y(t) = s(t)[1, 1]^T + n_2(t)[0, 1]^T,$$

see the figure below. Draw the first pair of canonical correlation vectors $w_x$ and $w_y$ in the figure.

14. A stochastic system has three states: 1, 2 and 3. The probability that the system moves from state 1 to state 2 is 1. The system moves from state 2 to states 1 or 3 with equal probabilities. From state 3, the system always moves to state 2. What is the probability that we (after a long time) find the system in state 1?
15. SVM: Draw the optimal hyper plane (decision boundary) and mark the support vectors in the following two cases:

a)

b)
16. The cocktail party problem: Assume we have access to three different mixes (linear combinations) $m_1$, $m_2$ and $m_3$ of three independent signals $s_1$, $s_2$ and $s_3$. We want to find these original signals. In order to solve the problem we will use canonical correlation analysis, CCA.

   a) What should we use as input data to CCA and what output data will get get? Explain the meaning of used variables and state their dimensions. (2p)

   b) How can we use the output from CCA to recreate the original signals? (2p)

   c) If BBS is used in this way in order to, for example, separate sound signals with the help of microphone recordings, certain problems may occur. Name one example of what could happen that would make the use of BBS more difficult. (1p)

17. You have $N$ number of $n$-dimensional training examples

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{pmatrix}, i = 1 \ldots N$$

Derive a way to find the first principal direction in this set of training data.
18. The figure shows a deterministic state model and a corresponding reward function. The states are enumerated from 1 to 5 and arrows represent actions denoted as "left" and "right". Numbers close to the arrows specify rewards.

![State Model Diagram](image)

**Figure 1: State model.**

a) Calculate the Q- and V-function for optimal policy as a function of $0 < \gamma < 1$. (3p)

b) Calculate the Q- and V-function for the $\epsilon$-greedy policy as a function of $0 < \gamma < 1$, where $\epsilon = \frac{1}{14}$. The $\epsilon$-greedy policy states that one chooses the action with the highest future reward with the probability $1 - \epsilon$. Any of the other alternative actions is chosen with the probability $\epsilon$. Observe that you have to use expectation values in the calculation of both the Q- and V-functions. This is due to the $\epsilon$-greedy strategy which adds a randomness to the system and the Q- and V-functions should represent the average expected reward. (2p)

19. Consider the following neural network.

![Neural Network Diagram](image)

tanh is used as activation function.

a) Derive an expression for the weight updates of both the layers. (3p)

b) Of what use is the activation function in the hidden layer? (1p)

c) Explain using words how the net is trained in practice and how it is used for classification. (1p)