Exam in
Neural Networks and Learning Systems
- TBMI26

Time: 2009-06-11 at 8-12
Teacher: Thord Andersson, Phone: 013-229449
Allowed additional material: Calculator, Tefyma, Beta, Physics handbook

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 (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 2009-06-25 at the latest. The exams will then be available at IMT.

GOOD LUCK!
Part 1

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

2. What is the meaning of Coover’s theorem in words?

3. What is defined by a kernel function?

4. What does the second principal component in PCA describe?
5. Mention a classifier that do not use an explicit discriminant function.

6. The signals \( \cos(2x) \) and \( \sin(3x) \) are mixed by the matrix

\[
\begin{pmatrix}
1 & 2 \\
2 & 4
\end{pmatrix}.
\]

What happens if this mixture is analysed using ICA?

7. What is optimized in canonical correlation analysis?
8. What is the difference between the “value function” \( V \) and the Q-function in Reinforcement learning?

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

10. What are the thread-like parts called that transmit the input signals in biological neurons?
11. If the number of samples is smaller than the dimensionality, the data covariance matrix becomes singular. Show that PCA can be performed by an eigen value decomposition of the inner-product matrix (the Gram matrix) of the samples.

12. After training an SVM, the resulting discriminant function $f(\varphi)$ is given by:

$$f(\varphi) = 4\varphi_1 + 9\varphi_2 + 4\varphi_3 - 16\varphi_4$$

The samples, $x$ have been mapped according to:

$$\varphi_1(x) = x_1^2$$
$$\varphi_2(x) = x_2^2$$
$$\varphi_3(x) = x_1 x_2$$
$$\varphi_4(x) = 1$$

Is any of the two samples $x = (1 \ 1)^T$ and $y = (1 \ -1)^T$ a support vector? Why/why not?
13. What class does X belong to according to a k-NN classifier with k=3? Why? What class does X belong to according to SVM? Draw the separating planes and support vectors. (Draw carefully so that it is clear if vectors are orthogonal or parallel etc.!) 

14. a) How many correlations with corresponding projection vectors will we find if we analyze the following multi-dimensional signals with CCA? (1p)

\[
x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
\]

\[
y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \\
\]

b) what is the relation between the different projection vectors that are found for \( x \), i.e. \( w_{x_1}, w_{x_2} \) etc? (1p)

15. Suppose you want to store a list of names and phone numbers in a Hopfield memory. How would you do that? What is the advantage of storing it in such a way?
16. You have been given \( N \) input-output signal pairs \( \{x, d\} \).

a) Derive an expression for the optimal weight vector that minimizes the square error \( \epsilon = E[(d - y)^2] \), where \( y \) is generated by a perceptron with a linear activation function. (3p)

b) Calculate the optimal weight vector given the input signals \( x_1 = \begin{bmatrix} 4 \\ 8 \end{bmatrix} \), \( x_2 = \begin{bmatrix} 2 \\ 6 \end{bmatrix} \), \( x_3 = \begin{bmatrix} 1 \\ 5 \end{bmatrix} \) and \( x_4 = \begin{bmatrix} 3 \\ 7 \end{bmatrix} \) and the desired output signals \( d_1 = 36, d_2 = 26, d_3 = 21 \) and \( d_4 = 31 \). (2p)

17. a) You will perform an update of a self-organizing network (SOM). The network consists of 4 units in a 1D topology. Use a learning factor of 0.5. The neighborhood function is \( h(i, j) = \exp(-\frac{d(i,j)^2}{2\sigma^2}) \), where \( \sigma \) is 1. The prototype vectors \( w \) before the update and input data \( x \) are shown in the figure below.

What are the weights after the update? (3p.)

b) How does a cost function, minimized by updates in (a), look like? Assume \( h(i, j) \) approaches 0. (2p.)
18. You do not have to normalize the vectors in this question.

    a) We have two multidimensional signals \( \mathbf{x} \) and \( \mathbf{y} \) according to:

\[
\mathbf{x} = \begin{pmatrix}
\sin(\omega t) + n_1 \\
\sin(\omega t) + n_2
\end{pmatrix}
\]

\[
\mathbf{y} = \begin{pmatrix}
\sin(\omega t) \\
\cos(\omega t)
\end{pmatrix}
\]

\( n_1 \) and \( n_2 \) are independent realizations of white normal noise with the same variance.

If we use CCA om these signals, we get two sets of correlations \( \rho \) and projection directions \( \mathbf{w}_x \) and \( \mathbf{w}_y \). What will \( \mathbf{w}_x \) and \( \mathbf{w}_y \) be for the first correlation? (2p)

b) Assume instead that we have

\[
\mathbf{x} = \begin{pmatrix}
\sin(\omega t + \phi) + n_1 \\
\sin(\omega t + \phi) + n_2
\end{pmatrix}
\]

\[
\mathbf{y} = \begin{pmatrix}
\sin(\omega t) \\
\cos(\omega t)
\end{pmatrix}
\]

Now, what will \( \mathbf{w}_x \) and \( \mathbf{w}_y \) be for the first correlation? (2p)

c) We have the same \( \mathbf{x} \) and \( \mathbf{y} \) as in part b. What will \( \rho \), \( \mathbf{w}_x \) and \( \mathbf{w}_y \) be for the second correlation? (1p)

19. The figure shows two different deterministic state models and the corresponding rewards. The states are enumerated from 1 to 6 and the arrows represent actions denoted as "forward" and "shortcut". The numbers close to the arrows states the corresponding awards.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig1.png}
\caption{The state models A and B.}
\end{figure}

    a) Describe the optimal policy for the system A and B respectively. (1p)
    b) Calculate the Q- and V-functions for the \( \epsilon \)-greedy-policy for system A as a function of \( \gamma \) (0 < \( \gamma \) < 1). (2p)
    c) Calculate the Q- and V-functions for the \( \epsilon \)-greedy-policy for system B as a function of \( \gamma \) (0 < \( \gamma \) < 1). (2p)

NB: The \( \epsilon \)-greedy-policy states that we will choose another way than the optimal with the probability 0 < \( \epsilon \) < 1.