# Hertentamen Neurale Netwerken

*20 August 1998*

**Grading:** In total you can get 100 points. You get 10 points free. Grade = number of points/10.

## Problem 1 (35 points)

***Perceptron & Linear Separability***

**a) (5p)** Let us consider two sets of points in $R^3$:

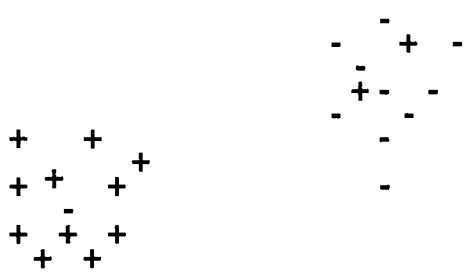
A={(0,0,0), (0,0,1), (0,1,0), (1,0,0)}

and

B={(1,1,1), (1,1,0), (1,0,1), (0,1,1)}.

Are these two sets lineary separable? Justify your answer (make a small drawing)

**b) (10p)** Two sets, each having 100 points, are 'almost' linary separable: it is enough to remove 3 points to make these sets separable (otherwise they are not separable):



Suppose that you have applied 2 algorithms to train a single perceptron to separate these two sets: the peceptron algorithm and the pocket algorithm. For each of these algorithms answer (with YES or NO) the following questions:

a) the network will stabilze (the weights will not be modified anymore)

b) the weights will 'oscilate' forever

c) the weights will tend to +/- infinity

d) the network will learn to separate both sets

e) the network will almost learn to separate both sets: there will be at most 3 misclassified points

**c) (20p)** Show that a sinlge layer perceptron with 35 input and 10 output units can learn to recognize ten bitmaps (7x5 = 35 pixels each) that represent digits 1, ..., 9, 0.
More precisely, let us assume that digints are represented by ten input vectors, D1, ..., D9, D0 (each of them of length 35) and ten output vectors: (1, 0, ..., 0), (0, 1, 0, ..., 0), ..., (0, ..., 1, 0), (0, ..., 0, 1) (each of length 10). Given these vectors find values of weights and biases of all the units that implement the desired mapping.

## Problem 2 (20 points)

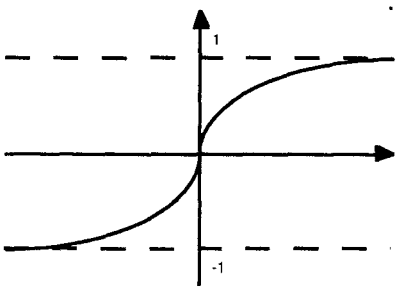
***Backpropagation***

**a) (4p)** Which function is minimized by the backpropagation algorithm? Define this function (write a formula).

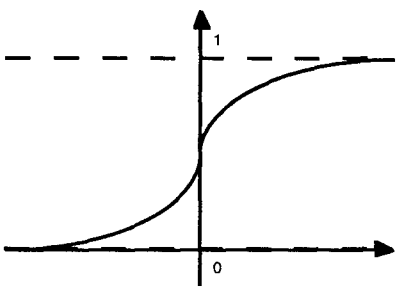
**b) (6p)** What do we assume about transfer functions? Which of the transfer functions given below can be used in backpropagation networks:

- logistic
- hyperbolic tangent
- threshold
- bounded linear
- linear
- sin

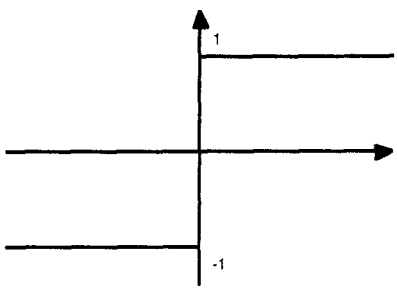
Justify your answer with one sentence.



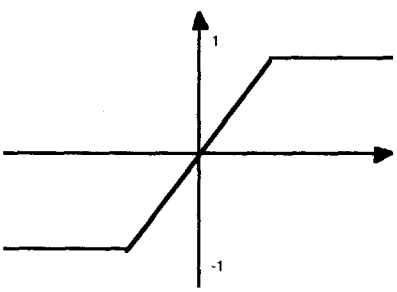

a) logistic



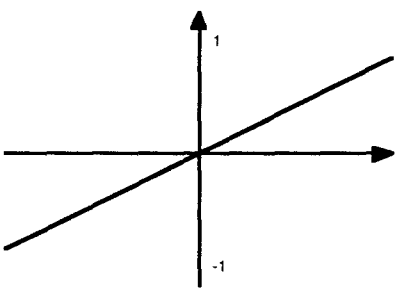

b) hyperbolic tangent



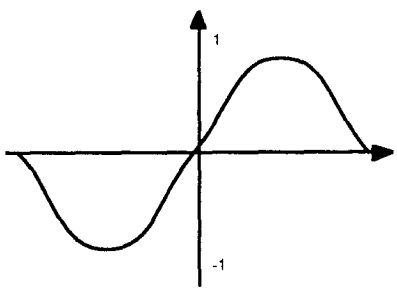

c) threshold




d) bounded linear




e) linear




f) sinus

**c) (4p)** It is a common practice that instead of using 0 and 1 as desired output values we use 0.1 and 0.9. What is the idea behind this practice?

**d) (6p)** What happens when the value of the learning rate ($\eta$) is not properly chosen? Consider 4 situations:

- $\eta$ is very small (and $\eta > 0$),
- $\eta$ is very big (and $\eta > 0$),
- $\eta$ is negative (e.g., $\eta = -1$),
- $\eta = 0$.

## Problem 3 (20 points)

The NetTalk system which was developed by Rosenberg and Sejnowski was used for converting words (represented by strings of characters) into phonems.

**a) (6p)** Describe the NetTalk system. What was the architecture of the network ? How were the data represented? Which training algorithm was used? What were the results?

**b) (14p)** Which of the following networks could be used instead of NetTalk and which could not? Briefly justify your answers.

a) single perceptron
b) multi-layer feedforward network
c) recurrent network
d) RBF network
e) SOM (Kohonen network)
f) Hopfield network
g) Cascade-Correlation algorithm

## Problem 4 (15 points)

**a) (10 p.)** Describe SOM (Kohonen network). What is the architecture of the network? How is it trained? How does it work? What are the most typical applications of such networks?

**b) (5 p.)** A well-known example of a SOM is a network that maps a 2-D picture of a cactus (or a square) into a 2-dimensional grid of nodes. Describe the following details of this example:

a) what is the training set?
b) how many input units are used?
c) how do we demonstrate the result of training?

**Remark.** If you have never heard about the NetTalk (it was not mentioned in the book of Mehrotra at al.) do the following assignment:

**a) (10p)** Describe the ART1 network and the training algorithm. What is the main difference between ART1 and ART2 networks?

**b) (5p)** Give an example of an application of the ART1 network.

**c) (5 p)** Comment on the role of the vigilance parameter. What would be the effect of making it decrease with time?