Hernani Marques (07-804-719) / Thomas Kägi (07-915-127)

NN / Task Sheet 1

Q1:
State the “five basics” of neural networks and illustrate them with an example.

(1) Existence of input and output layers of neurons, perhaps an hidden layer
(2) Some sort of back-propagation algorithm
(3) a learning form, e. g. supervised learning, semi-supervised learning, unsupervised learning
(4) There are training rules for the network and cost functions
(5) The neurons are connected (through synapses) and influence each other

Q2:
State three differences between biological neural networks and conventional computers.

(1) Biological systems are fault-tolerant
(2) Biological systems operate in an highly parallel fashion, whereas computer systems are usually serial
(3) Biological systems can generalize, whereas computers can't

Q3:
In artifical neurons, we use an abstract notion of weights, output, input and connections between neurons. Figure 1 shows a biological neuron. Point out the entities to which these four abstract terms correspond.

<u>Artifical NN</u>	<u>Biological NN</u>
Weights	Inhibitory and excitatory connections
Output connections	Axons
Input connections	Dendrites

Q4:
The NOR function is to be learned with the simple perceptron shown in figure 2. Complete the table below assuming a learning rate of $\eta = 0.6$, a binary threshold of $\Theta = 0$ and the perceptron learning rule $w_i(t + 1) = w_i(t) \pm \eta x_i(t)$. Note that weights are only updated if $o = \zeta$.

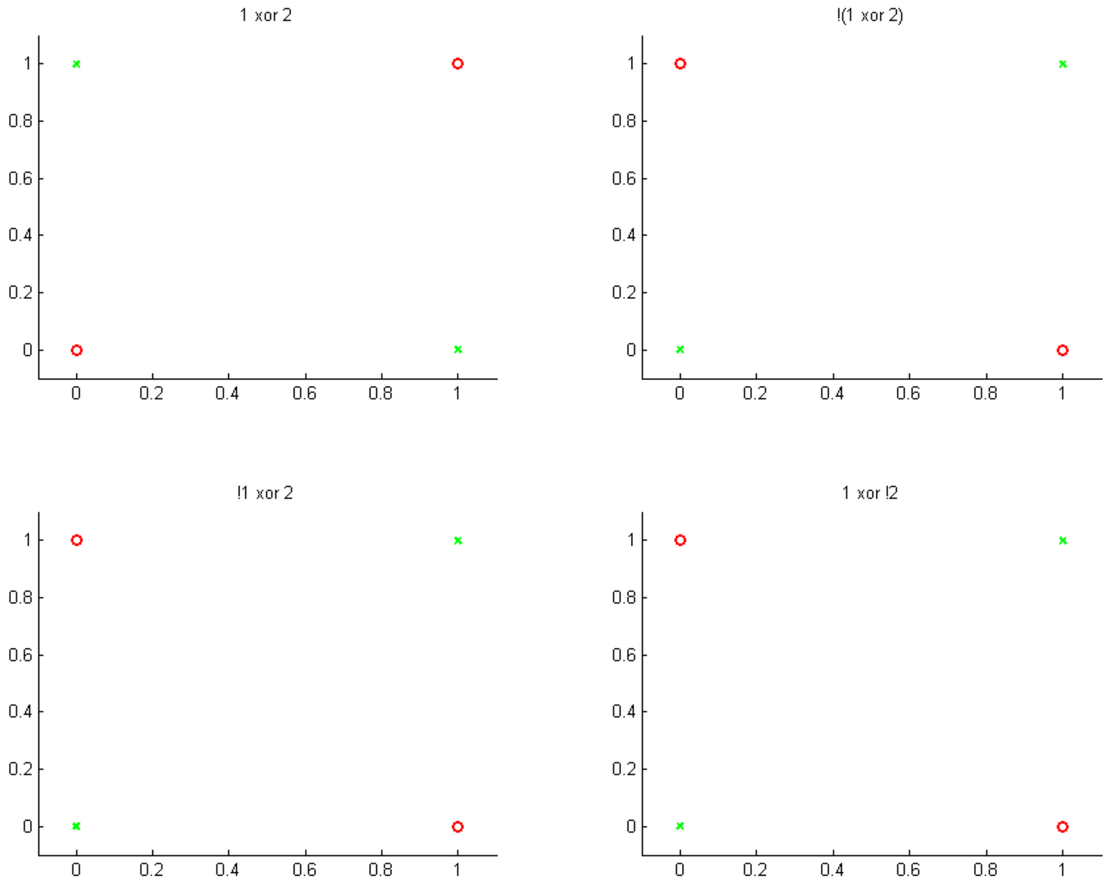
ξ_0	ξ_1	ξ_2	w0	w1	w2	$\sum w_i \xi_i$	o	ζ
1	0	0	-0.4	-0.5	0.5	-0.4	0	1
1	0	1	0.2	-0.5	0.5	0.7	1	0
1	1	0	-0.4	-0.5	-0.1	-0.9	0	0
1	1	1	-0.4	-0.5	-0.1	-1.0	0	0
1	0	0	-0.4	-0.5	-0.1	-0.4	0	1
1	0	1	0.2	-0.5	-0.1	0.1	1	0
1	1	0	-0.4	-0.5	-0.7	-0.9	0	0
1	1	1	-0.4	-0.5	-0.7	-1.6	0	0

1	0	0	-0.4	-0.5	-0.7	-0.4	0	1

Q5:
There are 16 different binary functions with two inputs and one output. Explain why this is.

There are two inputs with two possible values (binary), which result in 2^2 possible input patterns. And then there are two possible output values (binary) for each pattern, which means there are in total (2^2)^2 = 16 possible combinations, equivalent to functions.

Q6:
Out of all the above functions, draw the ones that are not linearly separable into coordinate systems. Use the two axes for the inputs and different symbols for the output and explain why they are not linearly separable.

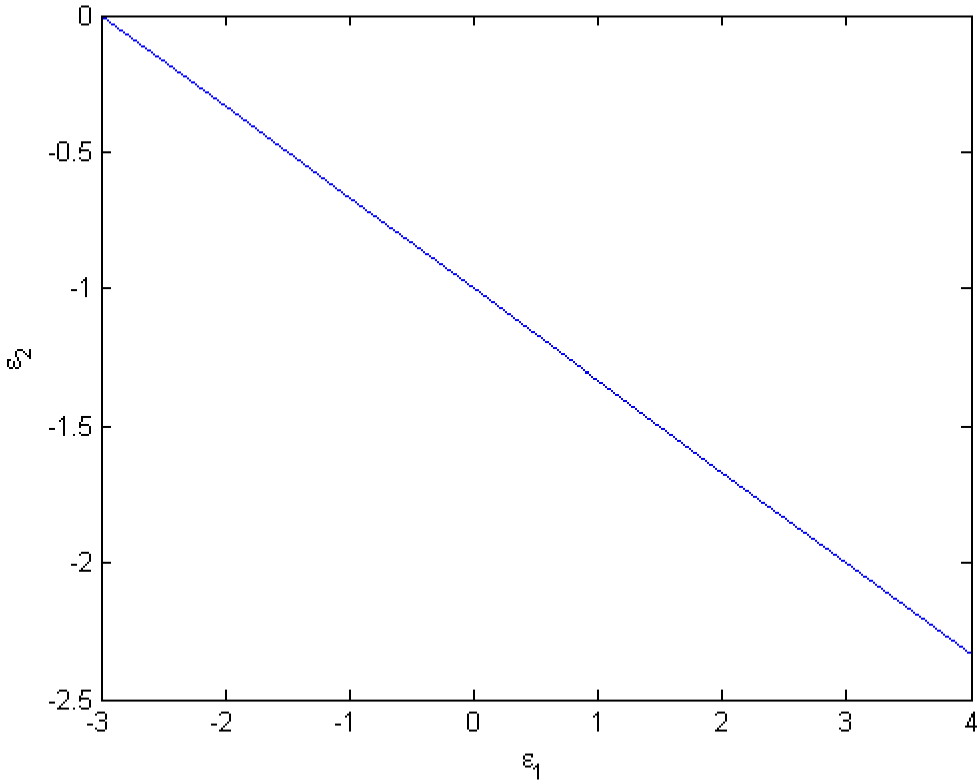


Green points symbolize 1 outputs, and red is 0.

Obvisouly no line can be drawn to separate the two colors, i. e. these binary functions are not linearly separable

Q7:
Given a perceptron with a weight vector wT = (w0 , w1 , w2) = (3, 1, 3) where node 0 is the bias node with ξ0 = 1, draw a sketch of the decision boundary in R2 and mark for which area the

perceptron outputs 0 and 1, respectively.



For the area above the line the perceptron outputs 1, for the area below it, it outputs 0.

Q8:
Consider a simple perceptron with 4 input nodes and 1 output node. ξ (i) is an input pattern and ζ (i) is the corresponding desired output. For the following two cases, state whether the patterns are linearly separable and/or linearly independent.

a)
ξ (1) = (1, 0, 1, 0)T, ζ (1) = 0
ξ (2) = (1, 0, 0, 0)T, ζ (1) = 1
ξ (3) = (0, 1, 0, 0)T, ζ (1) = 0
ξ (4) = (0, 0, 0, 0)T, ζ (1) = 1

b)
ξ (1) = (1, 0, 0, 0)T, ζ (1) = 0
ξ (2) = (0, 0, 1, 1)T, ζ (1) = 1
ξ (3) = (0, 1, 0, 0)T, ζ (1) = 0
ξ (4) = (0, 0, 0, 0)T, ζ (1) = 1

a) is linearly independent, b) isn't. But we cannot make statements about their linear separability, because we don't have enough pattern information. At least five patterns would be necessary.

Q9:

For the following statements, tell whether they hold true or not:

a) Adalines and Perceptrons use the same activation function.
False (script, page 24)

b) The learning rules for Perceptron, Adalines and MLP can be described by the generalized delta rule. Only the delta differs in the three rules.
False

c) Adalines use a threshold, whereas Perceptrons do not.
False (script, page 24)

d) Perceptrons adjust their weights only in the case of an error, Adalines adjust them always.
False

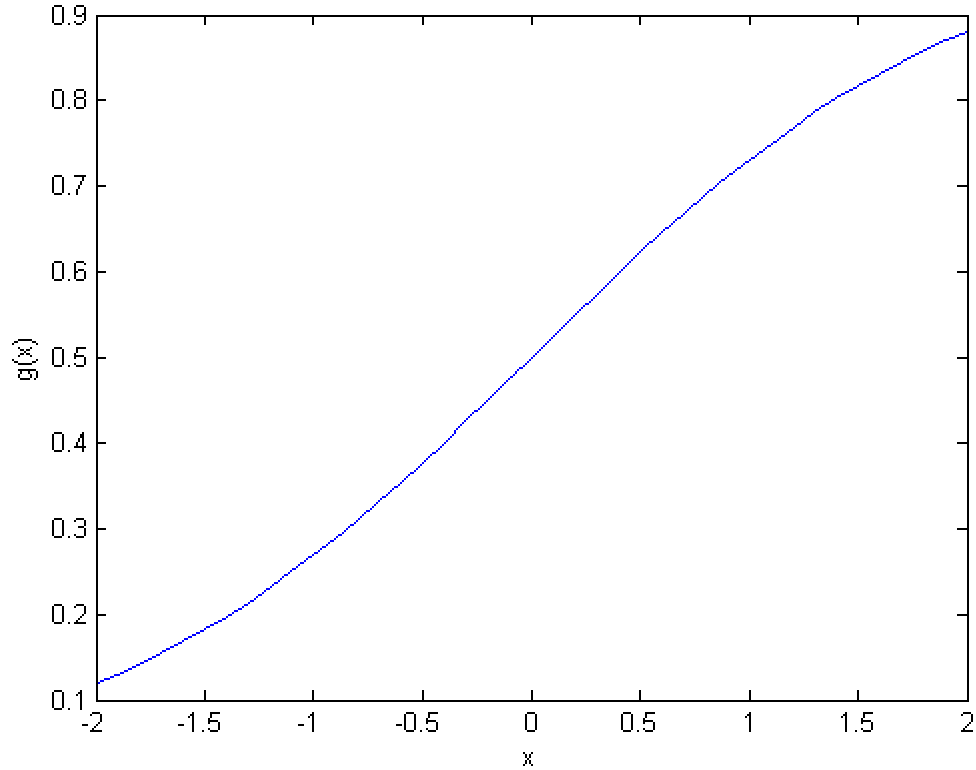
e) For the Perceptron there exists an explicit solution (without learning), if the classes are linearly independent.
False

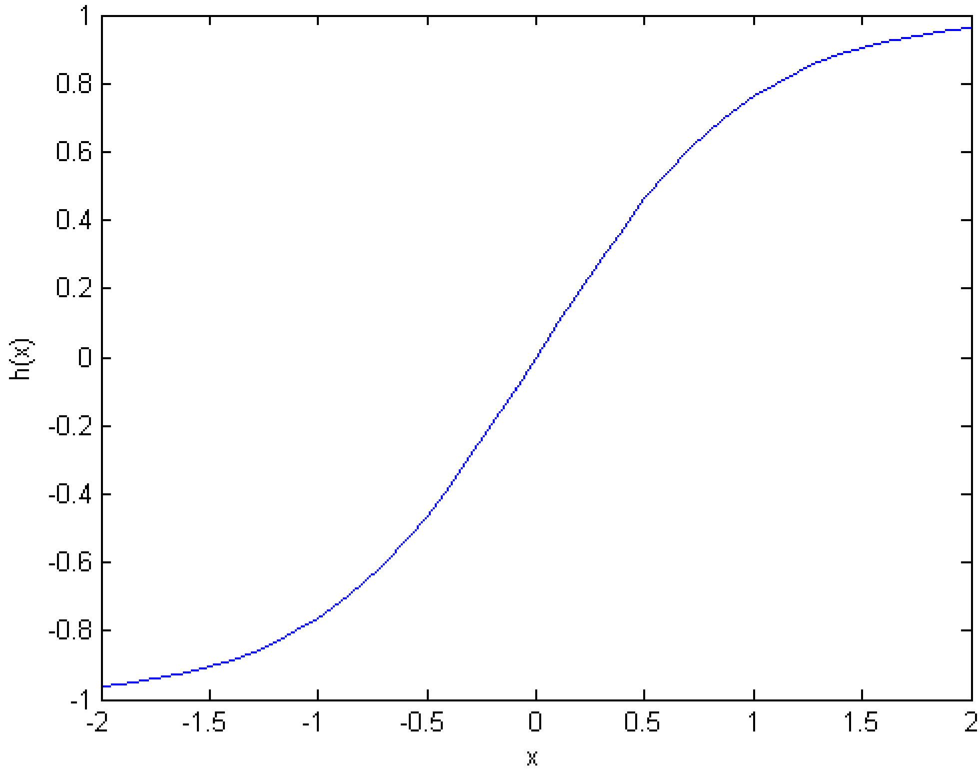
f) Perceptrons and Adalines can only learn linearly separable problems successfully.
True (script, page 26)

g) By adding more layers, Adalines can learn non-linear problems.
True (using Madalines, Multiple Adalines)

h) MLPs can learn non-linear problems successfully.
True (script, page 29)

Q10:
Plot the two functions





$g(x) = 1 / (1 + \exp(-x))$, $h(x) = \tanh(x)$,
compute their derivatives and verify the properties
$dg(x) / dx = g(x) \; (1 - g(x))$ and $dh(x) / dx = 1 - h(x)^2$

The derivatives are:

$$g'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$h'(x) = (\tanh(x))' = (\frac{\sinh(x)}{\cosh(x)})' = \frac{1}{\cosh(x)^2}$$

Proof of statet equalities:

$$g(x) \cdot (1-g(x)) = \left(\frac{1}{1+e^{x}}\right)\left(1-\frac{1}{1+e^{x}}\right) = \frac{e^{-x}}{(1+e^{-x})^2} = g'(x)$$

$$1-h(x) = 1-\frac{\sinh^2(x)}{\cosh^2(x)} = \frac{1}{\cosh^2(x)} = h'(x)$$

Q11:
A simple perceptron is used to learn patterns. The learning algorithm did not converge after 10^6 iterations. Can you claim that the patterns belong to classes that are not linearly separable? Justify your answer.

No, the claim would be false. Reasoning:

From the lecture we have the statement:
(i) If a problem is linearly separable, then an algorithm can be found that converges after a finite number of steps. ('=>')

In contrast, the above state meant does not strictly imply the opposite: if an algorithm doesn't converge that doesn't mean the problem isn't linearly separable.
('<=' does not hold. I.e: 'lin.Sep => convergence' does not imply 'no convergence => not lin. sep')

Also, we don't know the structure of the pattern. Hence the number 10^6 does not imply that the algorithm has already completed enough steps for convergence to occur.

See also script (as of page 22)

Q12:
The decision boundary of a simple perceptron is always a hyperplane. In the case of two inputs, it is a straight line in the input space R2 . Show mathematically why this is (for two inputs).

Perceptrons follow the decision rule:

$$\vec{\varepsilon}^{T} \cdot w \geqslant \Theta \rightarrow 1, 0 \, otherwise$$

in two dimensions we have

$$\vec{\varepsilon}^{T} = [\varepsilon_0 \varepsilon_1]$$

and the above equation takes the form

$$\varepsilon_0 \cdot w_0 + \varepsilon_1 \cdot w_1 = \Theta$$

on the boundary Θ, which is the equation for a hyperplane in R2, also known as the common line.