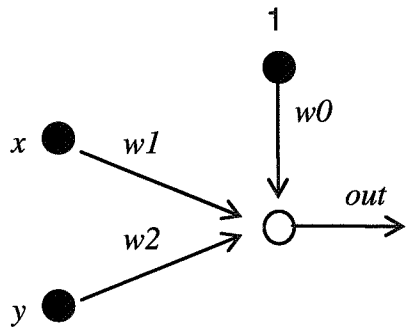
Exam Neural Networks

Wojtek Kowalczyk
20.10.2008

It is a "closed book" exam: you are not allowed to use any notes, books, etc. You may formulate your answers in Dutch or in English. For each problem you get some points; additionally you get 10 points for free. The final grade for this exam is the total number of points you get divided by 10. There are 5 problems.

1) A Simple Network (3+3+4+5+5+5 points)

Let us consider a single unit network with two inputs, one output, and the logistic sigmoid activation function $f(a)=1/(1+e^{-a})$:



a) Express the output of the network, *out*, as a function of five variables: x, y, w_0, w_1, w_2 (write a formula).

b) Express the error made by the network on the input (x,y) and the target output t as a function of x, y, t, w_0, w_1, w_2 (write a formula).

c) Describe the gradient descent minimization algorithm.

d) Derive the weight update rules for our network.

e) Can this network be trained to solve the XOR-problem? Justify your answer.

f) Now let us assume that instead of the logistic activation function we use the sinus function: $f(a)=sin(a)$. Can our network be now trained to solve the XOR-problem? Justify your answer.

In questions e) and f) we assume that the output of the network, *out*, is interpreted as 0 when $out < 0.5$, and 1 when $out \geq 0.5$.

2) Backpropagation (6*5 points)
Let us consider a multi-layer perceptron with 100 inputs, 10 hidden nodes in the first layer, 5 nodes in the second hidden layer and one output node. Let us assume that there are no bias nodes and that all the nodes use the logistic activation function.

a) Describe working of the network (forward pass).

b) Describe the backpropagation algorithm (include the formulation of the generalized delta rule).

c) How many weights are used by the network?

d) How many operations (additions, multiplications, calling the activation function) should be performed in order to produce an output on a single input vector?

e) How many operations are needed for a single step of the backpropagation algorithm (propagating errors backward and updating all weights)?

f) Is it true that during the training process the error (Mean Squared Error) on the (finite) training set cannot grow indefinitely? Justify your answer.

3) Thermometer representation of data (3+7 points)

a) Describe a "thermometer representation" of data.

b) Demonstrate that a network with a linear activation function may calculate a non-linear function of a single variable x, provided that values of x are represented with help of a thermometer representation.

4) The NETtalk System (5*3 points)
Describe in detail (as much as you can) the NETtalk system:

a) What was the purpose of the system?

b) What was used as a training set?

c) How were the inputs and outputs represented?

d) What was the network architecture?

e) What accuracies were achieved?

5) Winner-Take-All Networks (10 points)
Describe the architecture, working and training a Winner-Take-All (WTA) network. What are typical applications of these networks?