Exam Neural Networks

October 23, 2009

W. Kowalczyk
wojtek@cs.vu.nl

It is a "closed book" exam: you are not allowed to use any notes, books, etc. You may formulate your answers in Dutch or 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.

1) Bayes' Theorem (25 points)
Let us suppose that you have to develop, with help of Bayes' Theorem, a system that decides if an input image of a fruit is a banana (B) or an apple (A). The system has to use only one binary feature of input patterns: the color of the fruit, which may be either yellow (Y) or green (G), and no other colors are possible. The majority of bananas (70%) is yellow, while apples are usually green (90%). Moreover, it is known that apples are more frequent than bananas: only 30% of input patterns represent bananas, while the remaining 70% represent apples.

a) Formulate the Bayes' Theorem and describe how would you use it for solving this classification problem.

b) Suppose that your system sees a yellow fruit. What is the chance that it is an apple? Will your system label it as an apple? Why?

c) Suppose that your system sees a green fruit. What is the chance that it is an apple? Will your system label it as an apple? Why?

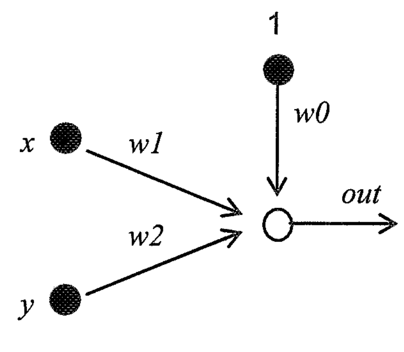
d) What is the accuracy of your system?

e) Is it possible to build a better classification system that uses as input the same single feature of input patterns (color, and nothing else) but has a better accuracy than implied by Bayes' Theorem? Briefly justify your answer.

f) Let us suppose that your system would be tested on a biased sample of fruits that consists of 1000 apples and 1000 bananas. What accuracy would be observed?

2) A Simple Network (25 points)

Let us consider a single unit network with two inputs, one output, and a non-standard activation function $f(a)=cos(a)$, see below:



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 w_0, w_1, w_2 (write a formula).

c) Describe the gradient descent minimization algorithm.

d) Derive the weight update rules for the network.

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

f) Can this network be 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 -1 when $out < 0$, and +1 when $out \geq 0$. Moreover, we assume that the logical values *true* and *false* are represented by +1 and -1, respectively:

x1	x2	OR(x1,x2)
-1	-1	-1
-1	+1	+1
+1	-1	+1
+1	+1	+1

x1	x2	XOR(x1,x2)
-1	-1	-1
-1	+1	+1
+1	-1	+1
+1	+1	-1

3) The NETtalk System (15 points)

Describe in detail (as much as you can) the NETtalk system:

a) What was the purpose of the system?
b) What were the train and test sets?
c) How were the inputs and outputs represented?
d) What was the network architecture?
e) What accuracies were achieved?

4) Thermometer representation of data (25 points)
A continuous variable x which takes values in [0, 1] is represented with help of a "5-level thermometer scale". In other words, each value of x is uniquely represented by 5 numbers from [0,1], $(d_1(x), d_2(x), d_3(x), d_4(x), d_5(x))$ in such a way that:

$$x=0.2*(d_1(x) + d_2(x) + d_3(x) + d_4(x) + d_5(x))$$

and if any $d_i(x)>0$ then all $d_j(x)$ with $j<i$ are equal to 1. For example:

x=0.00 is represented by (0, 0, 0, 0, 0),
x=0.15 is represented by (0.75, 0, 0, 0, 0),
x=0.25 is represented by (1, 0.25, 0, 0, 0),
x=0.47 is represented by (1, 1, 0.35, 0, 0).

Using this "thermometer representation" of x, a network with 5 input nodes and a single linear output node can be viewed as a device which computes a function $f(x)$, f: $[0,1] \rightarrow$ R, that is given by $f(x)=w_0+w_1*d_1(x)+...+w_5*d_5(x)$, where $w_0, ..., w_5$ denote weights of the network.

a) Make a plot of the function $d_2(x)$, for $0 \leq x \leq 1$.

b) Show that for a fixed set of weights $f(x)$ is a continuous, piecewise linear function of x.

c) Show that for a fixed training set $T=\{(x_i, t_i): i = 1..100, 0 \leq x_i \leq 1\}$ the mean squared error of the network is a polynomial of degree 2 in $w_0, w_1, ..., w_5$.

d) Suppose that our network is trained (MSE is minimized) on a training set that is shown in Figure A. Sketch a possible output of the trained network on the interval [0, 1].

e) Now suppose that our network is trained on an "incomplete and discontinuous" training set that is shown in Figure B. Sketch a possible output of the trained network on the interval [0, 1]. Describe the behavior of the network around the point $x=0.2$ and in the interval [0.4, 1].