# Exam Neural Networks

January 06, 2010

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 (20 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 feature of input patterns: the color of the fruit, which may be either yellow (Y), green (G), or red (R), and no other colors are possible. Bananas are usually yellow (80%), sometimes green (20%), and never red (0%). Apples, on the other hand, are usually red (60%), green (30%), and sometimes yellow (10%) . 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?

c) Suppose that your system sees a green fruit. What is the chance that it is a banana?

d) What is the accuracy of your system, i.e., what is the percentage of fruits that are correctly classified by your system?

Remark. If is not required that you perform calculations; it is sufficient to write down formulas, without finding their values.

**2) Perceptron and Linear Separability (10+5+5+10 points)**

a) Describe working of a single perceptron. Describe in detail the two learning algorithms: Perceptron and Pocket. What are the main convergence properties of these algorithms?

b) Is it correct that a single unit with a logistic sigmoid activation function may compute a non-linear decision boundary? Justify your answer!

c) Let us consider a training set that consists of 10 patterns that represent digits 0, 1, .., 9, each pattern being a vector of 256 of 0's and 1's that represents a 16x16 bit map image of a digit. Is it possible to train a single perceptron to separate the digit 7 from the remaining nine digits? Justify your answer.

d) Formulate the Cover Theorem which expresses the proportion of linearly separable dichotomies of size N in a d-dimensional space. (A plot of F(N,d) as a function of d/(N+1), for several values of d would do...).

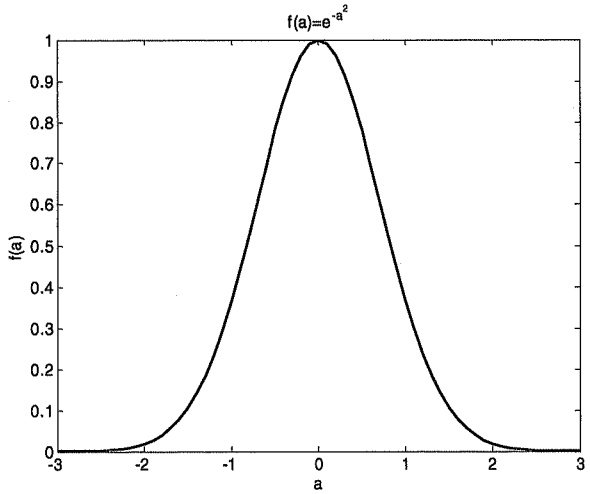
**3) Encoder networks and PCA networks (5 + 10 points)**

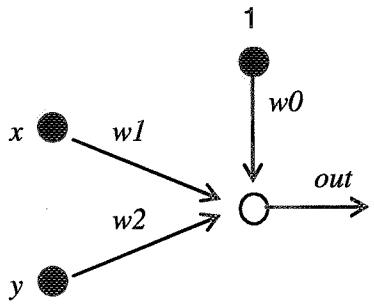
a) Describe the concept of an Encoder Network. What is the architecture of this network, what are the inputs and outputs, and what is this network supposed to learn?

b) Describe the concept of a PCA Network. Briefly describe one (or more) applications of PCA networks.

**4) A Simple Network (25 points (5+5+5+10))**

Let us consider a single unit network with two inputs, one output, and a non-standard activation function $f(a)=exp(-a^2)$, 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) Derive the weight update rules for the network (hint: $f'(a)=-2af(a)$).

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

In the last question we assume that the output of the network, *out*, is interpreted as -1 when $out < 0.5$, and +1 when $out \geq 0.5$. Moreover, we assume that the logical values *true* and *false* are represented by +1 and -1, respectively:

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