

## Neural Networks Exam

22 October 2015

This is a "closed book" exam: you are not allowed to use any notes, books, etc. You are allowed to use a **simple** calculator. Please read the questions carefully, formulate your answers clearly, and use either English or Dutch, grouping answers to the same question together (e.g. 1A-1C should not be interrupted by 3D). Ideally, your answers should be short and concise, focusing on the (sub)problems/questions listed. There are 90 marks you can get by addressing the problems below, and 10 marks will be given to you for free. Your final grade for this exam will be the total number of your marks divided by 10. Good luck!

### 1. Quick questions - short answers (for 40 marks overall)

**(A) (3 marks)** Name three features of the human brain that have contributed to the development of neural networks as an alternative to the traditional Von Neumann architecture.

**(B) (2+2 marks)** When are two classes considered linearly separable? Provide a graphical example of two linearly separable classes.

**(C) (3 marks)** Explain the principle of Maximum Likelihood to fit a given distribution towards a dataset.

**(D) (2+2 marks)** Express the function to update the weights of a neuron using the Adaline training rule, and explain the working of the update function.

**(E) (2+2 marks)** Express Cover's Theorem and explain how it has been used as a source of inspiration for more recent developments within neural networks.

**(F) (2+2 marks)** Name one of the major differences between the NEAT approach and HyperNEAT within the domain of neuroevolution, and explain the impact of this difference in terms of learning.

**(G) (2+4 marks)** What are the two approaches mentioned during the lecture to perform Principal Component Analysis? Explain for both approaches how they work.

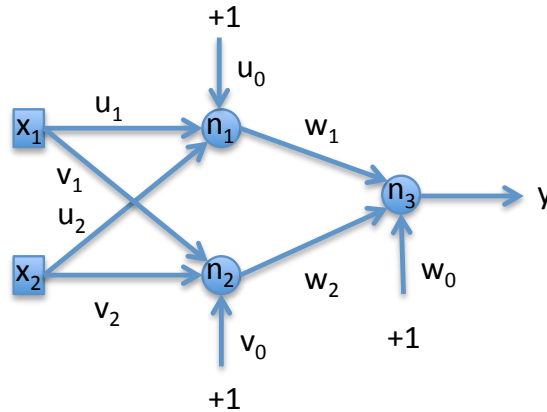
**(H) (3 marks)** Explain what the Kernel trick means in the context of Support Vector Machines.

**(I) (1+2 marks)** Name one of the optimization algorithms to speed up backpropagation that has been treated during the lecture and explain how it works.

**(J) (3+3 marks)** Write down the three principles of self-organization in the context of self-organization maps, and explain how the principles are reflected in self-organizing maps.

## 2. Multi-Layer Perceptron (24 marks, see breakdown below)

This assignment concerns a Multi-Layer Perceptron. The Figure below shows the layout of the network we are going to study. Here,  $x_1$  and  $x_2$  are inputs,  $n_1$  and  $n_2$  are hidden neurons in the first layer, and  $n_3$  is a neuron in the output layer. Furthermore, the Figure shows the weights and biases associated with each of the neurons. The value  $y$  is the output of neuron  $n_3$ .



For the assignments, assume the activation function is the identity function (i.e.  $\phi(a) = a$ ).

**(A) (4 marks)** Express the output value  $y$  as a function of the input  $x_1$  and  $x_2$ .

**(B) (4+2 marks)** Give the definition of the delta rule and explain briefly how the rule works.

**(C) (6 marks)** In the derivation of the delta rule, the following error function is assumed:

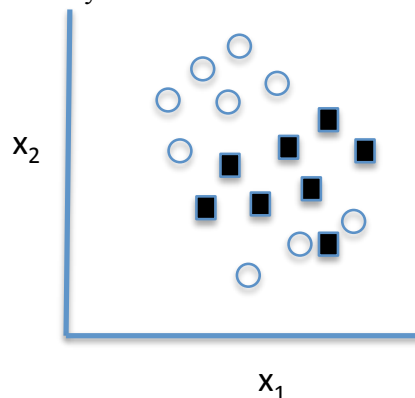
$$\sum_{i=1}^N (y_i - d_i)^2$$

Where  $N$  is the number of training examples. However, now we are going to assume a different function, namely:

$$\sum_{i=1}^N (y_i - d_i)^4$$

Assume gradient descent will be used. Derive how to update weight  $w_1$  using gradient descent and the new error function. For your derivation, you can omit the sum function (i.e. assume  $(y-d)^4$  as the error function).

**(D) (2+4 marks)** Look at the figure below. Here the circles and the squares represent two different classes. Would the network we are considering in this assignment be able to separate the two classes from each other? Provide arguments for your answer.



**(E) (1+1 marks)** Provide two stopping criteria for the back propagation algorithm that have been treated during the lecture.

### 3. Decision functions (26 marks; see breakdown below)

Let us consider a simple classification problem: predicting whether a person is a man or a woman based on the weight of that specific person. Assume that the attribute weight has five possible values: 1, 2, 3, 4, and 5 (representing weights below 60 kilograms, between 60 and 65, between 65 and 70, between 70 and 75 and above 75 respectively, but we will just stick to the 1-5 values). A total of 40 people (25 males and 15 females) are used as a training set. The table below shows the amount of people falling into each weight/sex category. For instance, for weight value 2 two male and ten females were counted.

weight	1	2	3	4	5
male	0	2	5	10	8
female	2	10	2	1	0

**(A) (2+4 marks)** Calculate the joint probability for each class/weight combination.

**(B) (2+4+2 marks)** Specify how the posterior probability is defined according to Bayes Theorem. Calculate the posterior probabilities for both classes for each possible value of the attribute weight (i.e.  $P(\text{male}|\text{weight}=1)$ ,  $P(\text{female}|\text{weight}=1)$ ,  $P(\text{male}|\text{weight}=2)$  etcetera). Draw a graph of these posterior probabilities.

**(C) (2+2+2 marks)** Give a mathematical definition of the optimal decision boundary. Use this formula to calculate the optimal decision boundary for the case described above and draw the decision boundary in the graph you have drawn under (A).

Now let us assume for the problem at hand that misclassifying a man as a woman is much more costly than misclassifying a woman as a man. More specifically, the cost of a misclassification of a woman as a man are 1, whereas a misclassification of a man as a woman costs 2.

**(D) (6 marks)** Recalculate the optimal decision boundary given the associated costs of misclassification. Clearly show how you come to your answer.