

Neural Networks Exam

23 October 2014

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, which will then be rounded to the nearest half. Good luck!

1. Quick questions - short answers (for 41 marks overall)

(A) (2 marks) What role did linear separability and the XOR problem play in the history of neural networks?

(B) (2+2 marks) Name two important differences between the architecture of the Von Neumann machine and the human brain.

(C) (2+2 marks) Explain the principle of maximum likelihood and explain how it relates to Bayes rule.

(D) (2+2 marks) Name two approaches within the field of neuro evolution that have been treated during the lecture. Briefly explain the difference between them.

(E) (2+2 marks) Explain the concept of regularization and provide an example of a regularizer for a neural network.

(F) (3+3 marks) Give the three principles of self-organization as they are used within self-organizing maps and explain for each of the principles how they are reflected in the rules that govern self-organizing maps.

(G) (1+1+2+2 marks) Name two approaches which can be used to adapt the learning rate for neural networks that have been discussed during the lecture. Explain for both of them how they work.

(H) (3 marks) Explain the main idea behind a gradient descent approach to minimize the error in a neural network, illustrate your explanation in a graphical manner.

(I) (3+1 marks) Give the function to be minimized in order to achieve the maximum margin between two classes within Support Vector Machines. Furthermore, explain the role points that are not support vectors play on the positioning of the decision boundary.

(J) (2+2 marks) What two approaches can be used to find principal components? Explain each of the two approaches briefly.

2. RBF networks (24 marks, see breakdown below)

This assignment addresses a Radial Basis Function network. Hereby, the network will be used for a classification task. To be more precise, the network will be used for the following classification problem (I represents the input for the network of dimension 2, O the desired output, the cases are numbered 1 to 4):

$$I_1 = \langle 0.9, 0.9 \rangle ; O_1 = +1$$

$$I_2 = \langle 0.9, 0.5 \rangle ; O_2 = -1$$

$$I_3 = \langle 0.1, 0.9 \rangle ; O_3 = -1$$

$$I_4 = \langle 0.1, 0.1 \rangle ; O_4 = +1$$

(A) (6 marks) In the first phase of the RBF approach that is deployed, a k-means clustering approach is used. In this case, k is set to 2 and the initial values of the centers are set to $t_1 = \langle 0.2, 0.5 \rangle$ and $t_2 = \langle 0.9, 0.4 \rangle$. Apply the k-means clustering approach and calculate the centers that result. Explain all intermediate steps.

Assume a basis function which is expressed as follows: $\phi(\|x-t\|) = e^{-\|x-t\|}$.

(B) (2+2 marks) What is the difference between a localized and a non-localized function? Is the function ϕ specified above localized or non-localized?

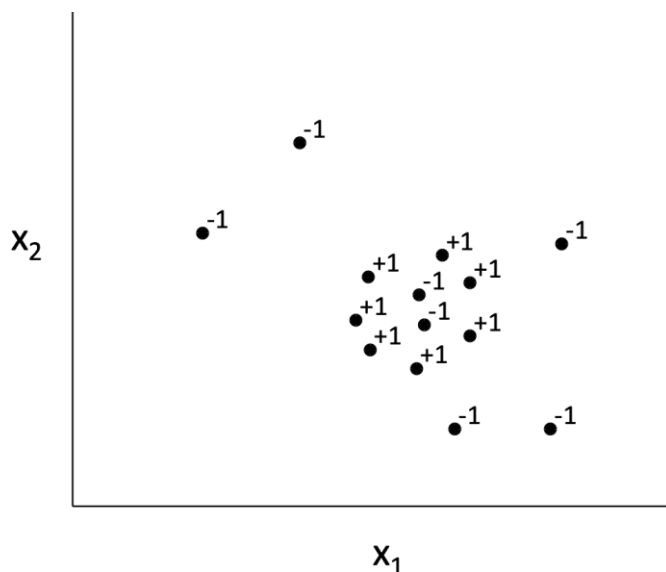
(C) (4 marks) Based on your answer of (A) calculate the output of the two neurons in the hidden layer for all four original data points given the function ϕ expressed above. If you were not able to answer (A) then assume the initial values of the centers given under (A).

(D) (2+2+2 marks) Is the output of the two hidden nodes linearly separable? Based on this answer, would you expect that the network is able to learn how to classify all cases correctly? Would a simple perceptron have been able to solve the initial problem?

(E) (2+2 marks) Is the RBF a supervised, unsupervised or hybrid learning approach? Provide a rationale for your answer.

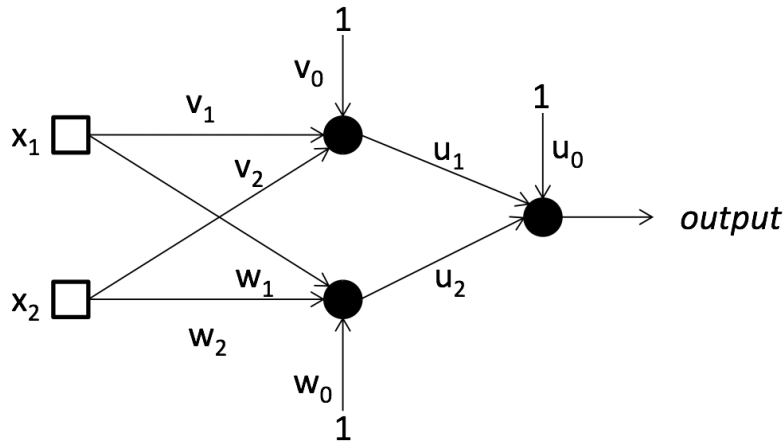
3. Multi-Layer Perceptrons (25 marks overall; see breakdown below)

Consider the dataset expressed below, where the class of each $\langle x_1, x_2 \rangle$ point is expressed next to it.



(A) (4 marks) Explain how many hidden layers you would select for the design of a multi-layer perceptron designed for solving this problem. In your explanation, take the characteristics of the problem as well as the different network setups into account.

Imagine that we are going to use the network shown below to solve the problem.



(B) (6 marks) Assume that all nodes have a linear decision function. Express the output of the network in terms of the inputs and the weights.

(C) (6 marks) Mathematically derive how weight u_1 should be updated under the assumption that the mean squared error function is used as the error function and gradient descent to determine the weight update.

(D) (2+3 marks) Give the general definition of the delta-rule, and explain how it can be used to update the weights in the network expressed above.

(E) (2+2 marks) Name two differences between RBF networks and multi-layer perceptrons in terms of the specification of the network.