Neural Networks Exam

24 October 2013

This is a "closed book" exam: you are not allowed to use any notes, books, etc. 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 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 25 marks overall)

(A) (3 marks) Why did ANN research become popular again in the 1980s?

(B) (6 marks) Name two approaches that can be used in Principal Component Analysis to determine the vector of principal components and briefly explain how they work.

(C) (6 marks) Explain the concept of a contextual map and explain how a Self-Organizing Map can be used to generate it.

(D) (4 marks) Is the first layer of an RBF network trained in a supervised manner? Justify your answer.

(E) (6 marks) Please describe two methods for weight initialization that have been treated during the lecture. If you do not know the formula for the not so obvious method, try to describe the main idea only.

2. Optical Digit Recognition (23 marks; see breakdown below)

Let us consider a problem of classifying images of handwritten digits into 10 classes: 0, 1, ..., 9, where each image is represented by a matrix of 16x16 = 256 numbers that represent pixel intensity. This problem can be addressed with 2 different methods:

a. a single layer perceptron (with 10 output nodes) with a threshold activation function.
b. a multi-layer perceptron with one hidden layer with 25 nodes and an output layer with 10 nodes (and all nodes using the logistic activation function).

(A) (6 marks) For each model, give the number of weights that should be trained.

(B) (3 marks) Describe the perceptron learning algorithm to find appropriate weights for the first network: the main idea and a pseudo-code (formulas are not required).

(C) (6 marks) Describe the backpropagation training procedure which can be used for training the second model, hereby express the formulas used to update the weights and explain them in words as well.

(D) (4 marks) Provide an alternative setup of a neural network which does not have 10 output nodes. You can use multiple networks if you want.

(E) (4 marks) Explain the difference between batch and incremental learning.

3. Support Vector Machines (24 marks overall; see breakdown below)

(A) (6 marks) Explain by means of a graphical example in what way a linear Support Vector Machine (SVM) tries to find a decision boundary between two linearly separable classes.

(B) (4 marks) What is the function that needs to be minimized in order to find the decision boundary you have specified under (A)? Furthermore, sketch the constraints under which this minimization should take place.

(C) (4 marks) What mathematical approach has been explained during the lecture to solve the minimization problem specified under (B)? Express the resulting equation for the linear SVM.

(D) (4 marks) How can Support Vector Machines be made suitable to separate non-linear problems? And how does this relate to Cover's theorem?

(E) (6 marks) Give three kernel functions that can be used for Support Vector Machines.

4. Decision functions (18 marks; see breakdown below)

Let us consider a simple classification problem: predicting whether a person has diabetes given the level of glucose in the blood. Assume that there are 4 possible values: 1, 2, 3 and 4 (representing actual glucose levels, the real values are not important for the exercise). A total of 30 people (10 diabetic and 20 non-diabetic) are used as a training set. The table below shows the amount of people falling into each glucose category and whether this person has diabetes or not. For instance, for glucose value 2 one person with diabetes was counted and six without diabetes.

glucose	1	2	3	4
diabetes	0	1	5	4
no diabetes	10	6	4	0

(A) (4 marks) Calculate the joint probability for each class/glucose combination.

(B) (6 marks) Calculate the posterior probabilities for both classes for each possible value of the attribute glucose (i.e. P(diabetes|glucose=1), P(no diabetes|glucose=1), P(diabetes|glucose=2) etcetera). Draw a graph of these posterior probabilities (i.e. the value for glucose versus the posterior probability for diabetes and no diabetes given that value).

(C) (4 marks) What would the optimal decision boundary be 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 diabetes patient as a not having diabetes is much more costly than misclassifying a person not having diabetes as a diabetic.

(D) (4 marks) Explain to what side the decision boundary you came up with under (C) would move given the scenario sketched above.