**Questions and Exercises**

**Course: Machine Learning and Applications**

**Chương 7: Neural Networks**

**7.1** Describe the difference between perceptron training rule and delta rule in training a perceptron.

*Mô tả sự khác biệt giữa quy tắc đào tạo perceptron và quy tắc delta trong đào tạo perceptron.*

**7.2** Describe the difference between *standard* *gradient descent* algorithm and *incremental gradient descent* algorithm in training a linear unit (perceptron).

**7.3** Consider a perceptron consisting of 2 inputs, one output and using a threshold function. Given the initial values for two weights *w1* = - 0.02, *w2* = 0.02 and the bias *w0* = 0.05, and the following training set with 4 patterns (each pattern consists of 2 attributes and one class label) as follows:

*x1 x2 y*

--------------------

0 0 0

0 1 1

1 0 1

1 1 1

a. Assume that the learning rate *η* is 0.25 and the used training algorithm is incremental gradient descent. After 4 training patterns are presented to the perceptron, what are the new values of *w0, w1, w2* ?

b. Which logic function does the perceptron represent?

c. Can incremental gradient descent algorithm for the case that perceptron is a linear unit can be used to determine the weights *λi* at the output nodes of RBF neural network?

**Ans:**

a) Notice that the input value corresponding to the bias w0 is always 1.

First, we present the pattern (0, 0, 0). The output value for this input pattern (0, 0, 0) is: *w0* × (1) + *w1*× 0 + *w2* ×0 = 0.05 × 1 = 0.05. This value is greater than 0, so the output of the perceptron is 1, which does not match with the desired value 0. The weight update rule is applied to adjust the weights and the weights become: (with *η* = 0.25):

*w0* = 0.05 + 0.25(0 -1) ×1 = -0.2

*w1* = -0.02 + 0.25(0 - 1) ×0 = - 0.02

*w2*= 0.02 + 0.25 (0 – 1) ×0 = 0.02

Next, we present the pattern (0, 1, 1). The output value for this input pattern (0, 1, 1) is: *w0*× (1) + *w1*× 0 + *w2* ×1 = - 0.2× 1 + 0 + 0.02 × 1 = - 0.18. This value is less than 0, so the output of the perceptron is 0, which does not match with the desired value 1. The weight update rule is applied to adjust the weights and the weights become:

*w0* = -0.2 + 0.25(1 – 0) ×1 = 0.05

*w1* = -0.02 + 0.25(1- 0) ×0 = - 0.02

*w2*= 0.02 + 0.25 (1 – 0) ×1 = 0.27

With the next pattern (1, 0, 1), the output value is not different from the desired output, so we do not have to adjust the weights.

With the next pattern (1, 1, 1), the output value is not different from the desired output, so we do not have to adjust the weights.

b) This perceptron represents the OR logic function.

c) The incremental gradient descent algorithm for the case that perceptron is a linear unit can be used to determine the weights *λi* at the output nodes of the RBF neural network sine the output units of RBF are also the linear units.

**7.4** Revise the backpropagation algorithm in order that this algorithm can works with the units in ANN using the transfer function *tanh* rather than the function *sigmoid*. The function *tanh* is defined as:

*tanh*(*x*) = (*ex* – *e-x*)/(*ex* + *e-x*)

Note: Modify the weight update rule at hidden layer and output layer. Notice that the differential of *tanh* has the property: *tanh’*(*x*) = 1 – [*tanh*(*x*)]2.

**7.5** Consider the ANN with the following configuration:



Initialize the weights and bias values as the following:

*w13 w14 w23 w24 w35 w45 w03 w04 w05*

-----------------------------------------------------------------------------

0.2 -0.3 0.4 0.1 -0.3 -0.2 -0.4 0.2 0.1

Assume that learning rate *η* = 0.9, the function *sigmoid* is used in the hidden layer and output layer, and backpropagation algorithm is used for training the ANN. If the training pattern <*x1*= **1**, *x2* = **0** và output = **1**> is presented to the training algorithm then how the weights and the bias values (*w03, w04, w05*) will be adjusted?

* 1. Explain the purpose of learning rate *η* and momentum constant *α* in the backpropagation algorithm for training ANN.
  2. State the advantages of ANN in classification.
  3. State the main weakness of ANN in classification.
  4. (True/false) ANN is more suitable with continuous data than categorical data.  
     *(Đúng/sai) ANN phù hợp với dữ liệu liên tục hơn là dữ liệu phân loại.*
  5. When determining the configuration of an ANN for classification, what do we base on to determine the number of nodes in the input layer and the output layer? And what method we can use to determine the number of nodes in the hidden layer ?   
       
     *Khi xác định cấu hình của một ANN để phân loại, chúng ta căn cứ vào đâu để xác định số nút ở lớp đầu vào và lớp đầu ra? Và chúng ta có thể sử dụng phương pháp nào để xác định số nút trong lớp ẩn?*

* 1. Describe step by step the process we apply to train RBF neural network?  
     *Mô tả từng bước quy trình chúng tôi áp dụng để đào tạo mạng lưới thần kinh RBF?*
  2. Describe the least-square method that can be used to estimate the output weights *λi* of a RBF neural network.  
       
     *Mô tả phương pháp bình phương nhỏ nhất có thể được sử dụng để ước tính trọng số đầu ra λi của mạng nơ-ron RBF.*
  3. Describe the main differences between MLP neural network and RBF neural network.

*Mô tả sự khác biệt chính giữa mạng thần kinh MLP và mạng thần kinh RBF.*

**Chương 8: Support Vector Machines**

**8.1** Describe some strong points of SVMs in comparison to MLP neural networks.

*Mô tả một số điểm mạnh của SVM so với mạng thần kinh MLP.*

**8.2** Explain the following terms: margin, maximum margin hyperplane, support hyperplane, support vector.

*Giải thích các thuật ngữ sau: lề, siêu phẳng lề tối đa, siêu phẳng hỗ trợ, vectơ hỗ trợ.*

**8.3** Prove that in the case of binary classification and the data is linearly separable, the margin is 1/||***w***|| where ***w*** is the weight vector of the separating hyperplane.

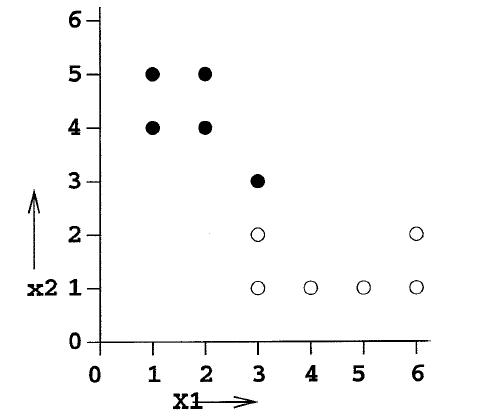
*Chứng minh rằng trong trường hợp phân lớp nhị phân và dữ liệu có thể phân tách tuyến tính thì lề là 1/||w|| trong đó w là vectơ trọng số của siêu phẳng phân tách.*

**8.4** In the case of binary classification and the data is linearly separable, describe the method to train SVM which is based on *constrained quadratic optimization problem*. Explain how to classify a test pattern using a trained linear SVM.

*Trong trường hợp phân loại nhị phân và dữ liệu có thể phân tách tuyến tính, hãy mô tả phương pháp huấn luyện SVM dựa trên bài toán tối ưu bậc hai có ràng buộc. Giải thích cách phân loại mẫu thử nghiệm bằng cách sử dụng SVM tuyến tính được đào tạo.*

**8.5** Given dataset with two real-valued attributes (x1 and x2) and one categorical output class as in the following figure. The positive points are shown as solid dots and the negative points are small circles.

Suppose you are using a linear SVM with no provision of noise. Draw three lines on the figure, showing the classification boundary and the two sides of the margin. Circle the support vectors.



**8.6** Given a training set with 9 patterns as follows:

|  | Attribute *x1* | Attribute *x2* | class |
| --- | --- | --- | --- |
| 1  2  3  4  5  6  7  8  9 | 0.5  1  0.5  1  1.5  4.5  5  4.5  5.5 | 3.0  3  2.5  2.5  2.5  1  1  0.5  0.5 | X  X  X  X  X  O  O  O  O |

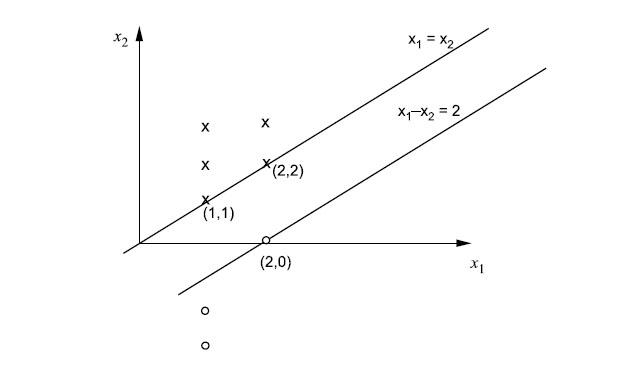
a) Prove that the patterns in class X satisfy the property *x1* – *x2* < 0 and the patterns in class O satisfy *x1* – *x2*> 0.

b) Show that the lines *x1* – *x2* = 0, 2*x1* – 2*x2* = -1 and *x1* – *x2* = 1 are the decision boundaries that can separate the two classes X and O. Find the weight vector ***w*** and the bias *b* for each such decision boundary.

c) Given the decision boundary with the equation *x1* – *x2* = 1. Show that the distance from any point of the form (*α,α*) where *α* is a real number, to the decision boundary is 1/sqrt(2).

**8.7** Given a dataset consisting of 8 patterns with two classes as in the following figure.

Use the training method for SVM to determine the weight vector ***w***, the bias *b* and the Lagrange multipliers *αi*, where *i* = 1,…8.



**8.8** Describe the difference between linear programming problem and quadratic programming problem.

**8.9** State the names of the two algorithms that can solve the quadratic programming problem arising in training a linear SVM for binary classification.

**8.10** Describe the purpose of kernel functions in SVM for the case that the data is not linearly separable.

*Mô tả mục đích của các hàm kernel trong SVM đối với trường hợp dữ liệu không thể phân tách tuyến tính.*

**8.11** State one typical weakness of SVM in classification.

*Nêu một điểm yếu điển hình của SVM trong phân loại.*

**8.12** Explaining of the slack variables *ξi* and the parameter *C* in SVM with soft margin.

*(Giải thích về biến chùng ξi và tham số C trong SVM với lề mềm.)*