Introduction to Neural Networks : Exercise Sheet 2

John A. Bullinaria - 2004

The syllabus and terminology for the Introduction to Neural Networks module have changed considerably over the years. The following questions are typical of what might come up in the exam this year. The percentages indicate what fraction of the two hour exam they correspond to.

Question 1 (Based on Question 2 in May 2001 Exam)

(a) The Back-Propagation (BP) algorithm is often used for training feed-forward neural networks. Why do we need to calculate the gradient in the BP algorithm? **[5%]**

(b) When a BP algorithm is used, the error function must be differentiable. Why? **[2%]**

(c) In a feed-forward neural network trained by BP, what is back-propagated, and from where to where? **[3%]**

(d) In a feed-forward network using a sigmoidal (logistic) activation function, why might one avoid setting target output values 0.0 and 1.0, and prefer to use 0.1 and 0.9 (or even 0.2 and 0.8). **[5%]**

Question 2 (Based on Question 2 in Jan 1999 Exam)

(a) Explain what the following equation, and each symbol in it, represents:

$$E_{sse}(w_{ij}) = \tfrac{1}{2}\sum_p \sum_j \left(targ_j^p - out_j(in_i^p)\right)^2 \qquad \textbf{[6\%]}$$

(b) Explain what "gradient descent learning" means. **[6%]**

(c) What are local minima, and what problems do they cause. **[3%]**

(d) How can problems with local minima be avoided? **[3%]**

Question 3 (Based on Question 1 in Jan 2000 Exam)

(a) Explain why we prefer to use the logistic (sigmoid) function as the activation function in back-propagation networks. **[5%]**

(b) What does the learning rate do in back-propagation training? What may happen when we use a very large learning rate, and when we use a very small learning rate? **[5%]**

(c) Can a back-propagation training algorithm always find a set of weights which minimise the training error for a given feed-forward neural network? Explain the ideas behind your answer. **[5%]**

(d) What role does momentum play in back-propagation training? **[5%]**

Question 4

A typical weight update equation for artificial neural network training is:

$$\Delta w_{kl} = \eta \sum_{p} (targ_l - out_l).f'(\sum_{n} in_n w_{nl}).in_k$$

Describe in words what the various symbols, parameters and variables in it refer to, and how one might go about deriving such an equation. **[10%]**

Question 5 (Based on Question 2 in August 2003 Resit Exam)

(a) The basic equation for gradient descent training is

$$\Delta w_{kl} = -\eta \frac{\partial E(\{w_{ij}\})}{\partial w_{kl}}$$

Describe what the symbols in it refer to and how it can be used to derive a training algorithm for a Multi-Layer Perceptron (MLP). **[8%]**

(b) Explain the distinction between *Batch Training* and *On-line Training*. Give one advantage for each of them. **[4%]**

(c) Suppose you are using an MLP for classification. Justify a particular choice for the output activation function. **[4%]**

(d) Under what circumstances might you want to use MLP output targets of 0.1 and 0.9 rather than 0 and 1? Explain why. **[4%]**

Question 6 (Based on Question 2 in May 2002 Exam)

(a) Describe how the basic "Back-Propagation Learning Algorithm" for Multi-Layer Perceptron (MLP) networks is related to gradient descent learning. **[10%]**

(b) What are local minima, and why are they a problem? How might we improve our chances of finding the global minimum? **[5%]**

(c) Explain the purpose of the "momentum term" that is often included in the Back-Propagation Learning Algorithm. **[7%]**

(d) Outline what is meant by a line search and why we might want to use one. Explain what the "Conjugate Gradient" algorithm is, and which features of it result in improved speed of learning. **[8%]**