Introduction to Neural Networks : Exercise Sheet 4

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 6 in May 2001 Exam)

Discuss the similarities and differences between back-propagation (BP) and radial basis function (RBF) networks. **[15%]**

Question 2 (Based on Question 3 in May 2002 Exam)

(a) A company has collected a large amount of data in the form of pairs of input and output vectors, and wants you to build a system that will predict the outputs for new inputs. Design an appropriate Radial Basis Function (RBF) network for them. Explain what will be computed at each network layer. **[10%]**

(b) Explain how you will determine the input to hidden layer weights/parameters. **[5%]**

(c) Explain how you will determine the hidden to output layer weights/parameters. **[5%]**

(d) What is the main advantage your Radial Basis Function (RBF) network will have over a Multi-Layer Perceptron (MLP) network carrying out the same function? **[5%]**

Question 3 (Based on Question 3 in August 2002 Resit Exam)

(a) Two equations used in the context of Radial Basis Function (RBF) mappings are:

$$y_k(\mathbf{x}) = \sum_{j=0}^{M} w_{kj}\phi_j(\mathbf{x}) \qquad \phi_j(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}-\boldsymbol{\mu}_j\|^2}{2\sigma_j^2}\right)$$

Explain what each of the symbols in them mean, and what it is that is being computed by them. **[8%]**

(b) Show how the RBF mapping can be represented in the form of a network. **[6%]**

(b) How does one train such a network? **[10%]**

(d) What factors will affect our choice of the number of hidden nodes we use in such a network? **[6%]**

Question 4 (Based on Question 3 in August 2003 Resit Exam)

(a) Describe the architecture of a Radial Basis Function network with D input units and K output units, and explain what is computed at each layer. **[6%]**

(b) How would you use a set of training data to determine the weights/parameters of such a network? **[8%]**

(c) What is a Validation set? How would you use one to choose an appropriate number of hidden units for an RBF network? **[6%]**

Question 5 (Based on Question 5 in May 2001 Exam)

Ensemble averaging is a simple technique for constructing committee machines. A practical training strategy that is often used is to:

1. Start a set of expert neural networks from different initial random weights.
2. Over-train each of them, i.e. let them over-fit the training data.
3. Take linear combinations of the expert neural network outputs.

Since over-training causes over-fitting, doesn't this cause over-fitting of the committee machine? So, why do we over-train the individual expert neural networks? **[10%]**

Question 6 (Based on Question 4 in August 2002 Resit Exam)

(a) What are committee machines, and why might one want to use them? **[6%]**

(b) Committee machines can be classified as having 'static' or 'dynamic' structures. Explain what the distinguishing feature is, and give an example of each. **[4%]**

(c) Describe the 'mixtures of experts' committee machine, and outline its advantages over the 'ensemble averaging' approach. **[10%]**

Question 7 (Based on Question 4 in May 2003 Exam)

(a) Explain in general terms what is meant by the term "Committee Machine", and why such a thing might be useful. **[7%]**

(b) Describe the architecture of a "Mixtures of Experts" system. Outline how you would go about training such a system. **[7%]**

(c) Discuss how a "Mixtures of Experts" system could be used to automatically generate a modular neural network architecture. **[6%]**