Introduction to Neural Networks : Exercise Sheet 3

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 2004 Exam)

(a) In the context of feed-forward neural networks, explain what the following equation, and each symbol in it, represents:

$$E_{sse}(\{w_{kl}\}) = \frac{1}{2}\sum_p \sum_j \left(targ_j^p - out_j(in_i^p)\right)^2$$ **[5%]**

(b) Describe the basic ideas underlying *Gradient Descent Learning Algorithms*, and how the above equation would be used for such training. **[5%]**

(c) Outline how, in general, you could estimate the generalization ability of a trained neural network. **[5%]**

(d) How could you modify the above equation to improve the generalization ability of your trained neural network? **[5%]**

Question 2 (Based on Question 4 in May 2001 Exam)

(a) Why do we use early stopping in back-propagation training? **[5%]**

(b) What are the major advantages of early stopping over alternative procedures? **[5%]**

(c) What is k-fold cross validation? How is it relevant to early stopping? **[5%]**

Question 3 (Based on Question 2 in August 2002 Resit Exam)

(a) Explain how a *gradient descent algorithm* can be used to train Multi-Layer Perceptrons (MLPs). **[10%]**

(b) Outline the principal techniques one can use to speed up the training when using such algorithms. **[10%]**

(c) Outline the principal techniques for improving the generalization ability of networks trained in that way. **[10%]**

Question 4

(a) Explain in words what the following equation signifies:

$$\mathcal{E}_D\left[\left(\mathcal{E}[y\,|\,x_i]-net(x_i,W,D)\right)^2\right]$$
$$=\left(\mathcal{E}_D\left[net(x_i,W,D)\right]-\mathcal{E}[y\,|\,x_i]\right)^2+\mathcal{E}_D\left[\left(net(x_i,W,D)-\mathcal{E}_D\left[net(x_i,W,D)\right]\right)^2\right]$$
$$= \qquad (\text{bias})^2 \qquad + \qquad (\text{variance})$$

[5%]

(b) Using function approximation by a multi-layer perceptron as an example, describe the extreme cases of overfitting and underfitting within this framework. **[4%]**

(c) Describe three things we might vary in an attempt to improve the generalization ability of a multi-layer perceptron. **[6%]**

(d) Explain how either a validation set or cross-validation can be used to assist us in improving generalization. **[5%]**

Question 5 (Based on Question 2 in May 2003 Exam)

(a) In the context of a Multi-Layer Perceptron (MLP) training, explain carefully the relation between *Weight Decay* and cost function *Regularization*. **[5%]**

(b) Why might one expect weight decay to help improve the generalization ability of an MLP? **[5%]**

(c) Describe how you could use *Cross-Validation* to choose the appropriate level of weight decay for a given classification problem. **[5%]**

(d) Briefly outline two other approaches, apart from weight decay, that you could expect to improve the generalization ability of an MLP. **[5%]**

Question 6

Why might one want to add noise to the training data when training a neural network? Should the noise be added to the inputs, the target outputs, or both? Suggest a procedure for determining what level of noise is appropriate. **[12%]**

Question 7

What is weight decay in feed-forward network training? Why is it useful? Explain why we should not also decay the thresholds/biases. How can one determine what rate of weight decay is appropriate? **[12%]**