Neural Computation: Exercise Sheet 6
John A. Bullinaria - 2015

The following questions are of the kind that may come up in the exam this year. They are designed to help you monitor your progress – try to answer the questions without your notes, and then use your notes to check whether your answers are correct. The percentages indicate the corresponding fraction of a 1.5 hour exam.

Question 1

(a) What are Committee Machines, and why might one want to use them? [7%]

(b) Committee machines can be classified as having either static or dynamic structures. Explain what is the distinguishing feature, and give an example of each. [7%]

(c) Describe the Mixtures of Experts committee machine, and outline its advantages over the Ensemble Averaging approach. [11%]

Question 2

(a) Outline how a consideration of ensemble averages over all possible training sets can be used to derive the standard Bias + Variance Decomposition of expected generalization performance. (Detailed mathematical derivations are not required.) [14%]

(b) Ensemble Averaging is a simple technique for constructing Committee Machines, and a practical training strategy for them is to:

1. Start a set of individual neural networks from different initial random weights,
2. Over-train each of them, i.e. let them over-fit the training data, and
3. Compute outputs as averages over the individual neural network outputs.

Since over-training generally results in poor generalization, explain why this approach can be expected to result in improved generalization performance. [11%]

Question 3

(a) Explain in general terms what is meant by the term Committee Machine, and suggest what computational advantages such an approach might offer. [7%]

(b) Describe the architecture of a Mixtures of Experts system. Outline how you would go about training such a system. How does this depend on whether the task is of a regression or classification type? [11%]

(c) Discuss how a Mixtures of Experts system could be used to automatically generate a modular neural network architecture. [7%]
**Question 4**

(a) *Ensemble Averaging* and *Boosting Machines* are said to be *Committee Machines with a Static Structure*. Explain what that means. [8%]

(b) Outline the key ideas underlying *Boosting*. [8%]

(c) Describe briefly the three different ways of implementing Boosting generally referred to as *filtering, sub-sampling* and *re-weighting*. [9%]

**Question 5**

(a) Explain what is meant by the terms *mean*, *standard deviation*, and *standard error*. How are the *standard deviation* and *standard error* related? [8%]

(b) Explain what is meant by the terms *median* and *quartiles*, and suggest when these quantities may be more useful than the mean and standard deviation. [5%]

(c) What is meant if someone says that the difference in performance between two models is *statistically significant*? Describe how a *t test* and can be used in this regard. What conditions need to be satisfied for the results of a *t test* to be reliable? [8%]

(d) Is it possible for a difference in performance between two models to be statistically significant, but still not be important? Explain why. [4%]

**Question 6**

(a) Explain how a *validation set* or *cross-validation* can be used to determine the optimal learning or regularization parameters for a given neural network model. Include a description of the kinds of things that are typically optimized by this process, and any differences between classification and regression type problems. [12%]

(b) Outline how an *Evolutionary Algorithm* can be used to automate such an optimization process for a neural network application. Take care to specify the key stages and operators involved in such algorithms. [13%]

**Question 7**

(a) Explain how ideas from the evolution of real brains can be used to generate improved neural network models. [9%]

(b) Describe the kinds of problems that are likely to be encountered if an evolutionary approach is used to learn neural network weights for realistic applications. [8%]

(c) Outline how an evolutionary approach can be used to investigate *one* of the following: (i) The best neural network learning parameters for a given task, (ii) The evolution of brain modularity, or (iii) Learning as a factor in Life History Evolution. [8%]