Tik-61.261 Principles of Neural Computing

Raivio, Venna

Exercise 9 24.3.2004

1. One of the important matters to be considered in the design of a MLP network is its capability to generalize. Generalization means that the network does not give a good response only to the learning samples but also to more general samples. Good generalization can be obtained only if the number of the free parameters of the network is kept reasonable. As a thumb rule, the number of training samples should be at least five times the number of parameters. If there are less training samples than parameters, the network easily overlearns – it handles perfectly the training samples but gives arbitrary responses to all the other samples.

   A MLP is used for a classification task in which the samples are divided into five classes. The input vectors of the network consist of ten features and the size of the training set is 800. How many hidden units there can be at most according to the rule given above?

2. Consider the steepest descent method, $\Delta \mathbf{w}(n) = -\eta \mathbf{g}(n)$, reproduced in formula (4.120) and earlier in Chapter 3 (Haykin). How could you determine the learning-rate parameter η so that it minimizes the cost function $\mathcal{E}_{av}(\mathbf{w})$ as much as possible?

3. Suppose that we have in the interpolation problem described in Section 5.3 (Haykin) more observation points than RBF basis functions. Derive now the best approximative solution to the interpolation problem in the least-squares error sense.