Tik-61.261 Principles of Neural Computing

Raivio, Venna

Exercise 7

1. In section 4.6 (part 5, Haykin pp. 181) it is mentioned that the inputs should be normalized to accelerate the convergence of the back-propagation learning process by preprocessing them as follows: 1) their mean should be close to zero, 2) the input variables should be uncorrelated, and 3) the covariances of the decorrelated inputs should be approximately equal.

   (a) Devise a method based on principal component analysis performing these steps.

   (b) Is the proposed method unique?

2. A continuous function $h(x)$ can be approximated with a step function in the closed interval $x \in [a, b]$ as illustrated in Figure 1.

   (a) Show how a single column, that is of height $h(x_i)$ in the interval $x \in (x_i - \Delta x/2, x_i + \Delta x/2)$ and zero elsewhere, can be constructed with a two-layer MLP. Use two hidden units and the sign function as the activation function. The activation function of the output unit is taken to be linear.

   (b) Design a two-layer MLP consisting of such simple sub-networks which approximates function $h(x)$ with a precision determined by the width and the number of the columns.

   (c) How does the approximation change if tanh is used instead of sign as an activation function in the hidden layer?

3. A MLP is used for a classification task. The number of classes is C and the classes are denoted with $\omega_1, \ldots, \omega_C$. Both the input vector $\mathbf{x}$ and the corresponding class are random variables, and they are assumed to have a joint probability distribution $p(\mathbf{x}, \omega)$. Assume that we have so many training samples that the back-propagation algorithm minimizes the following expectation value:

$$E\left(\sum_{i=1}^{C}[y_i(\mathbf{x}) - t_i]^2\right),$$

   where $y_i(\mathbf{x})$ is the actual response of the ith output neuron and t_i is the desired response.

   (a) Show that the theoretical solution of the minimization problem is

$$y_i(\mathbf{x}) = E(t_i|x).$$

   (b) Show that if $t_i = 1$ when $\mathbf{x}$ belongs to class ω_i and $t_i = 0$ otherwise, the theoretical solution can be written

$$y_i(\mathbf{x}) = P(\omega_i|\mathbf{x})$$

   which is the optimal solution in a Bayesian sense.

   (c) Sometimes the number of the output neurons is chosen to be less than the number of classes. The classes can be then coded with a binary code. For example in the case of 8 classes and 3 output neurons, the desired output for class ω_1 is $[0, 0, 0]^T$, for class ω_2 it is $[0, 0, 1]$ and so on. What is the theoretical solution in such a case?

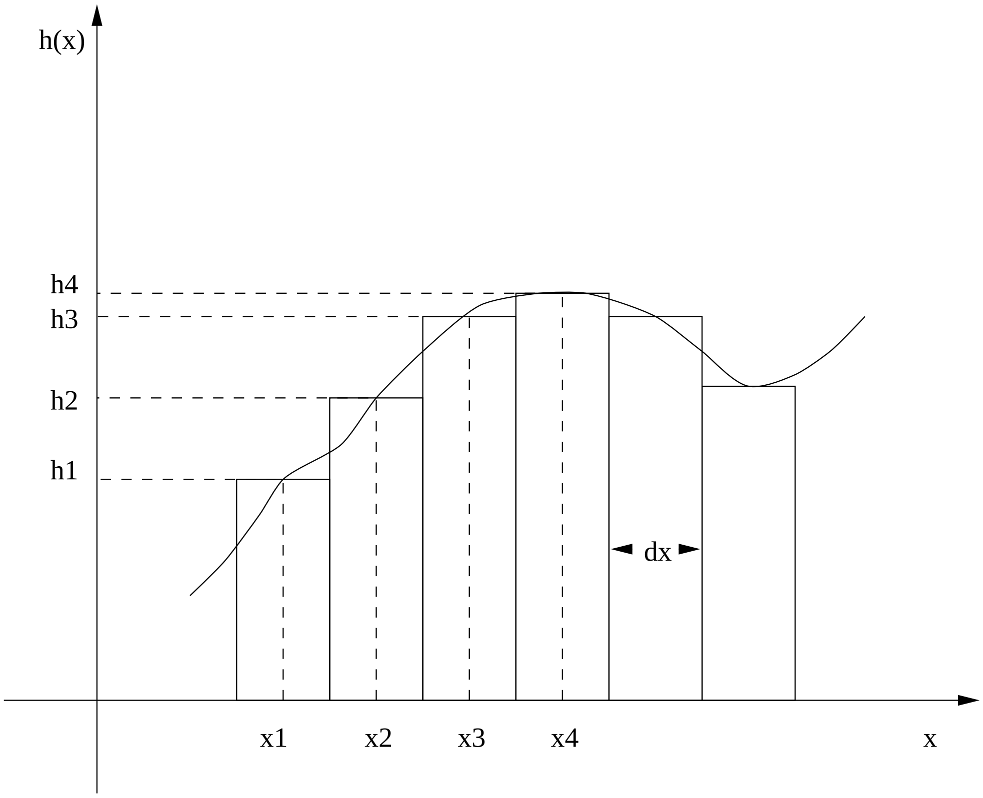

Figure 1: Function approximation with a step function.