T-61.261 Principles of Neural Computing
Examination 13.5.2005

Raivio

1. Answer briefly to the following questions or items (using a few lines or sentences only for each question):
   (a) What is the main deficiency of a single-layer perceptron?
   (b) What learning vector quantization (LVQ) is doing?
   (c) Why regularization is often needed in radial-basis function (RBF) networks?
   (d) How a MLP (multilayer perceptron) network trained using the back-propagation algorithm can be applied to unsupervised learning (for example to data compression)?
   (e) What kind of neighborhood function is typically used in training the self-organizing map (SOM)?
   (f) How a recurrent neural network differs from a customary feedforward network?

2. What heuristic methods do you know for making the back-propagation algorithm to perform better? - It suffices to explain briefly each method.

3. Consider the following interpolation problem which constitutes the mathematical background for radial-basis function (RBF) networks: Find a function $F(\mathbf{x})$ of the form

$$F(\mathbf{x}) = \sum_{i=1}^{M} w_i \varphi(\| \mathbf{x} - \mathbf{t}_i \|)$$

satisfying the conditions

$$F(\mathbf{x}_j) = d_j, \quad j = 1, 2, \dots, N.$$

Here $\mathbf{x}_j$, $j = 1, 2, \dots, N$, are known training input vectors, and the scalar numbers d_j are the respective outputs (desired responses). The vectors $\mathbf{t}_i$, $i = 1, 2, \dots, M$, are the centers of the radial basis functions $\varphi(\| \mathbf{x} - \mathbf{t}_i \|)$, and w_i, $i = 1, 2, \dots, M$, are the respective scalar weights.

   (a) Express the problem in matrix-vector form (using a suitable notation).
   (b) Assume first that there are as many ($N = M$) centers $\mathbf{t}_i$ as training pairs $(\mathbf{x}_j, d_j)$. Find the optimal weights w_i, $i = 1, 2, \dots, M$.
   (c) Assume now that there are more training pairs as centers, so that $M < N$. Find the best approximate solution to the interpolation problem in the least-squares error sense.

4. Consider the cost function

$$\mathcal{E}(\mathbf{w}) = 2w_1^2 - w_1 w_2 + 3w_2^2 + 7$$

where w_1 and w_2 are the components of the 2-dimensional parameter vector $\mathbf{w}$.

   (a) Which value of the parameter vector $\mathbf{w}$ minimizes the cost function $\mathcal{E}(\mathbf{w})$?
   (b) Construct a steepest descent algorithm for minimizing the cost function.
   (c) Construct Newton's algorithm for minimizing the cost function.
   (d) How quickly Newton's algorithm converges?