T-61.5130 Machine Learning and Neural Networks
Examination 14th December 2009/Karhunen

You are allowed to have in the examination a collection of mathematical formulas, but not any of the teaching material. (Voit vastata tenttiin myös suomeksi.)

1. Answer briefly (using a few lines) to the following questions or items:
   (a) What for is Oja's rule used in neural computing?
   (b) For which purpose is weight decay used?
   (c) Explain briefly ϵ-insensitive cost function.
   (d) Which are the two main criteria used for measuring non-Gaussianity of a random variable or signal?
   (e) Explain briefly how LVQ (learning vector quantization) differs from SOM (self-organizing map)?
   (f) Explain briefly what is NARX model.

2. The neural network shown in Figure 1 on the reverse side of this sheet has the activation function $\varphi(\cdot)$, input vector $\mathbf{x} = (x_1, x_2, x_3)^T$, weights k_1, k_2, k_3, k, w_1, w_2 and w_3, and biases b_1 and b_2.
   (a) Present the general formula for the output y of the network.
   (b) In the N-bit parity problem each component of the N-dimensional input vectors $\mathbf{x}$ is a binary number 0 or 1. The correct value of the parity function is $y = 0$, if the sum of the components of the vector $\mathbf{x}$ is even, and $y = 1$ if the sum is an odd number. The network shown in the figure can be used for solving the parity problem by choosing as the activation function the step function

   $$\varphi(u) = \begin{cases} 1 & , u > 0 \\ 0 & , u \leq 0 \end{cases}$$

   Examine whether the network of the figure solves the 3-bit parity problem correctly, when the parameters of the network are chosen to have the values $w_1 = w_2 = w_3 = 1$, $k_1 = k_2 = k_3 = 1$, $k = -2$, $b_1 = -0.5$, and $b_2 = -1.5$.

3. Compare on a general level the similarities and differences between multilayer perceptron (MLP) networks and radial-basis function (RBF) networks.

4. Consider the cost function

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

   where $\mathbf{w} = [w_1, w_2]^T$ is a two-dimensional weight (parameter) vector.
   (a) Find the optimal value $\mathbf{w}^*$ of the weight vector minimizing the cost function $\mathcal{E}(\mathbf{w})$.
   (b) Construct steepest descent algorithm for minimizing the cost function $\mathcal{E}(\mathbf{w})$.
   (c) Construct Newton's optimization algorithm for minimizing the cost function $\mathcal{E}(\mathbf{w})$.
   (d) How fast does Newton's optimization algorithm converge?

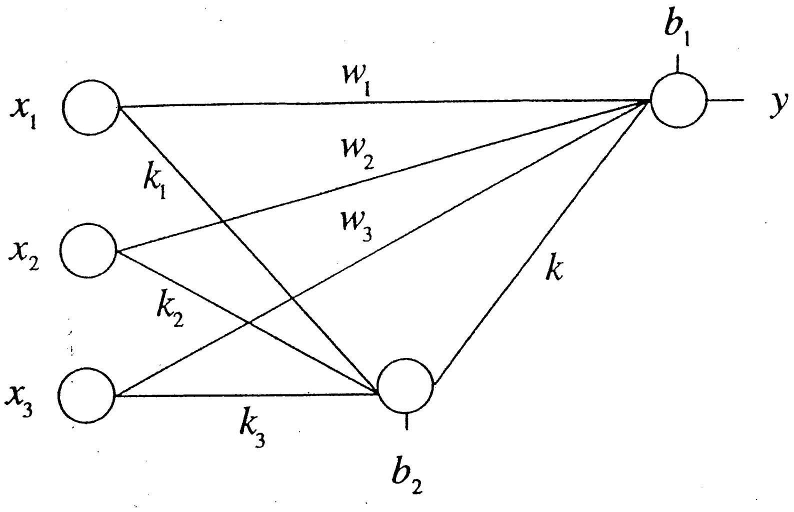

FIGURE 1