CS-E4810 Machine Learning and Neural Networks
Examination 16th December 2016/Karhunen

(Voit vastata tenttiin myös suomeksi.)

1. Answer briefly (using a few lines) to the following questions or items:
   (a) How does updating with a momentum term differ from the corresponding standard updating rule?
   (b) How is Jacobian matrix defined?
   (c) What means curse of dimensionality?
   (d) In which neural networks method one can use multiquadratic and inverse multiquadratic functions?
   (e) Which are the two main criteria for measuring non-Gaussianity?
   (f) Explain briefly what is NARX model.

2. Consider the general linear model for modeling a scalar variable

$$y : y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x})$$

   where $\mathbf{x}$ is data vector, $\mathbf{w}$ is M-dimensional weight vector with elements $w_0, w_1, ..., w_{M-1}$, and the $\phi_j(\mathbf{x})$, $j = 1, ..., M-1$ are some basis functions which can be nonlinear. Often $\phi_0(\mathbf{x}) = 1$ is the dummy 'basis function' corresponding to the bias term w_0. You have at your disposal N input-output training pairs $(t_i, \mathbf{x}_i)$. Model the dependence between input vector $\mathbf{x}$ and scalar output t using the general linear model above. Fit the model to the training data using the least-squares method with the added weight decay regularizer $0.5\lambda \mathbf{w}^T \mathbf{w}$.

3. Explain what are classification and regression trees (CART), and their advantages and drawbacks. How are random forests built from them?

4. Consider a supervised learning problem in which the output is scalar y and the desired response is d. Assume that we have trained for solving this problem two different neural networks whose outputs are respectively y_1 and y_2. Assume further that y_1 and y_2 are unbiased, and the noise term is neglected as is often done. Then the mean-square errors of the outputs y_1 and y_2 equal to their variances σ_1^2 and σ_2^2, respectively.

   Consider now the weighted output of the two networks

$$y = \alpha y_1 + (1 - \alpha) y_2$$

   where the weight α satisfies $0 \leq \alpha \leq 1$.

   (a) Is the weighted output y unbiased?
   (b) What is the mean-square error of y when y_1 and y_2 are assumed to be statistically independent of each other?
   (c) Find the value of α that minimizes the mean-square error of y.