1. Follow the example from the lecture (see Backpropagation Algorithm) and find the effect of changing $\omega_2(2, 3)$ on the output $X_3(1)$. Assume that we have the tangensoidal activation function with $\beta = 1$, a single neuron in the output layer, three neurons in single hidden layer and two fan-in neurons in the input layer. Of course, for each layer we assume additional “dummy neuron” with signal $X_0(0) = X_1(0) = X_2(0) = 1$.

2. Show that (following the above problem) for the $j$th neuron (here $0 \leq j \leq m$) in the first hidden layer (for network with one neuron in the output layer and $n$ neurons in the input layer) the incoming weight adjustments for $\Delta \omega_2(k, j)$ and $\Delta \omega_2(l, j)$ satisfy

$$
\Delta \omega_2(k, j) = \eta X_1(k) \delta_2(j) \quad \text{and} \quad \Delta \omega_2(l, j) = \eta X_1(l) \delta_2(j),
$$

where $\eta$ is a learning parameter. Assume here sigmoidal activation function for all network neurons with $\alpha = 1$. This result can be generalized on the other activation functions subject to minor modification. E.g. infer such modification for the tangensoidal activation function upon inspecting the solution to the problem 1. Compare your results with the sigmoidal activation function case. Both questions 1 and 2 refer to the Back Propagation Algorithm and its explicit structure ready for implementation.

3. How many linear dichotomies can we construct first for two points and then for four points positioned on the plane? Assume that the points are in the so-called general position. Show by example that the Function Counting Theorem does not hold for the points which are not in the general position.

4. We say that the sequence of $f_n(x)$ is convergent uniformly to $f(x)$ over the interval $I = [a, b]$ (denoted as $\lim_n f_n(x) = f(x)$) if

$$
\lim_n \left( \sup_{a \leq x \leq b} |f_n(x) - f(x)| \right) = 0.
$$

And we say that $f_n(x)$ is convergent to $f(x)$ point-wisely over $I$ if for any fixed $x_0 \in I$ we have

$$
\lim_n f_n(x_0) = f(x).
$$

Find the point-wise limit over $[0, 1]$ of the sequence of functions $f_n(x) = x^n$. Illustrate this on the diagram. Is it a uniform convergence? Do the same for $f_n(x) = x/n$ (the sequence of linear functions) and $g_n(x) = 1/n$ (the sequence of constant functions) first over the set of all real numbers $\mathbb{R}$ and then over the interval $[0, 1]$. This problem is to understand the approximation property of the neural net and what approximation and “closeness” of two results (if they are functions) could mean. Note that the uniform convergence of $f_n$ to $f$ over $I$ implies a point-wise convergence of $f_n(x_0)$ to $f(x_0)$ for each $x_0 \in I$ but not vice
versa. Note that the uniform convergence guarantees (a theorem in calculus) that for all $f_n(x)$ continuous the limit $f$ is also a continuous function. This may not happen with the point-wise convergence. See the examples specified in this problems.

Assoc. Prof. Ryszard Kozera