2 - Backpropagation with the Hyperbolic Tangent Function

Revise the backpropagation algorithm for two layers (see slide 34 of Lecture 2014-11-14) so that it operates on units using the hyperbolic tangent (squashing) function in place of the sigmoid function (see the remark on slide 28 of Lecture 2014-11-14). That is, the output o of a single unit is defined by

$$o = \tanh(net) = \tanh(\vec{w} \cdot \vec{x})$$

Give the weight update rule for output layer weights and hidden layer weights.

First we need to find a derivative of $o(x)$ function to use it in update procedure:

$$\frac{do}{dx} = \frac{d}{dx}\left(\frac{e^{2x}-1}{e^{2x}+1}\right) = \frac{(e^{2x}+1)\frac{d}{dx}(e^{2x}-1) - (e^{2x}-1)\frac{d}{dx}\ (e^{2x}+1)}{(e^{2x}+1)^2} =$$

$$= \frac{2e^{2x}(e^{2x}+1) - (e^{2x}-1) * 2e^{2x}}{(e^{2x}+1)^2} = \frac{2e^{2x}(e^{2x}+1-e^{2x}+1)}{(e^{2x}+1)^2} = \frac{4e^{2x}}{(e^{2x}+1)^2} =$$

$$= \left(\frac{2}{e^{x}+e^{-x}}\right)^2 = \operatorname{sech}(x)^2.$$

Having this derivative we can formulate update rules for output and hidden layer weights:

a) $\delta_k \leftarrow \frac{do}{dx} * (t_k - o_k) = \left(\frac{2}{e^{net}+e^{-net}}\right)^2 * (t_k - o_k)$

b) $\delta_h \leftarrow \frac{do}{dx} * \sum_{k \in outputs} w_{h,k}\delta_k = \left(\frac{2}{e^{net}+e^{-net}}\right)^2 * \sum_{k \in outputs} w_{h,k}\delta_k$

Having corresponding δ value for each output and hidden weight we can formulate next weight update rule:

$$w_{i,j} \leftarrow w_{i,j} + n\delta_i x_{i,j}$$