2 Artificial Intelligence (SBH)

This question is about neural networks. We consider initially multilayer perceptrons with nodes of the following kind.

\[
\sum_{i=0}^{n} w_i z_i \sigma(a) = \sigma(a)
\]

\(z_0 = 1\)
\(w_0\)
\(z_1\)
\(w_1\)
\(z_2\)
\(w_2\)
\(\sum_{i=0}^{n} w_i z_i \rightarrow \sigma \rightarrow z = \sigma(a)\)
\(z_n\)
\(w_n\)

(a) Derive an expression for the gradient \(\frac{\partial E_i(w)}{\partial w_j}\) for weight \(w_j\) in an output node when \(E_i(w)\) is the error for the \(i\)th example

\[E_i(w) = \frac{1}{2}(y_i - h(w; x_i))^2,\]

\(h(w; x_i)\) is the output of the complete network for the \(i\)th example, and \(\sigma(a) = a\). You need only derive the expression for the output node. [3 marks]

(b) Derive an expression for the gradient \(\frac{\partial E_i(w)}{\partial w_j}\) for weight \(w_j\) in an output node when \(\sigma(a) = 1/(1 + \exp(-a))\) and the error for the \(i\)th example is

\[E_i(w) = -y_i \log h(w; x_i) + (1 - y_i) \log(1 - h(w; x_i)).\]

You may use the fact that \(d\sigma(a)/da = \sigma(a)(1 - \sigma(a))\). You need only derive the expression for the output node. [7 marks]

(c) In the standard backpropagation algorithm the central quantity of interest for each node \(N\) is \(\delta = \partial E_i(w)/\partial a\). It is proposed that, instead of using nodes in the form presented above, we introduce functions \(\phi_i\) and construct multilayer networks from nodes that compute \(z = \sigma(a)\) where

\[a = \sum_{i=0}^{n} w_i \phi_i(z),\]

Here, \(z^T = [z_0 \ z_1 \ \cdots \ z_n]\) and the functions \(\phi_i\) are fixed, having no further parameters. A multilayer perceptron is constructed from nodes of this kind. Give a detailed, general derivation of the formula for computing \(\delta\) for a non-output node \(N\) in this network, assuming you know the values of \(\delta\) for the nodes connected to the output of \(N\). [10 marks]