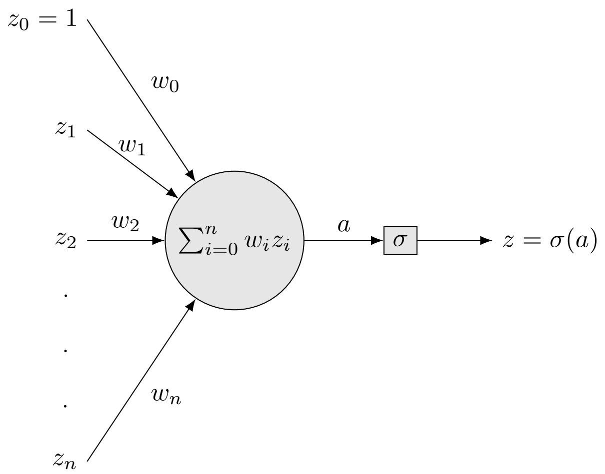
2 Artificial Intelligence (SBH)

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



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

$$E_i(\mathbf{w}) = \frac{1}{2}(y_i - h(\mathbf{w}; \mathbf{x}_i))^2,$$

$h(\mathbf{w}; \mathbf{x}_i)$ is the output of the complete network for the ith example, and $\sigma(a) = a$. You need only derive the expression for the output node. [3 marks]

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

$$E_i(\mathbf{w}) = -y_i \log h(\mathbf{w}; \mathbf{x}_i) + (1 - y_i)\log(1 - h(\mathbf{w}; \mathbf{x}_i)).$$

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 $\delta = \partial E_i(\mathbf{w})/\partial a$. It is proposed that, instead of using nodes in the form presented above, we introduce functions ϕ_i and construct multilayer networks from nodes that compute $z = \sigma(a)$ where

$$a = \sum_{i=0}^{n} w_i \phi_i(\mathbf{z}).$$

Here, $\mathbf{z}^T = \begin{bmatrix} z_0 & z_1 & \cdots & z_n \end{bmatrix}$ and the functions ϕ_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 δ for a *non-output* node N in this network, assuming you know the values of δ for the nodes connected to the output of N. [10 marks]