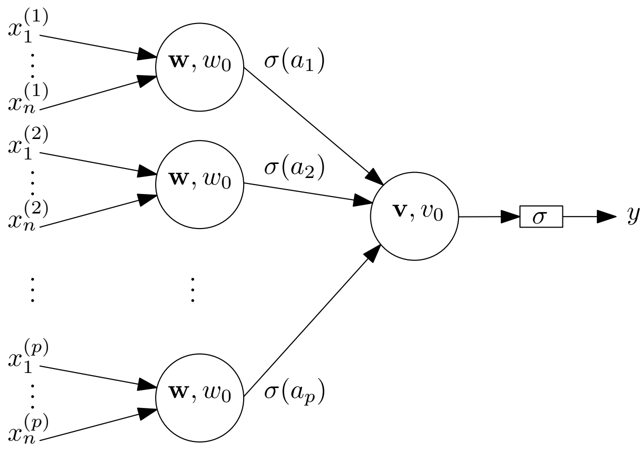
2 Artificial Intelligence (sbh11)

This question addresses a variation on the usual *multilayer perceptron*.



The input vector is divided into p groups each with n elements. Let $x_i^{(j)}$ be the ith element in the jth group and let $\mathbf{x}^{(j)} = \begin{pmatrix} x_1^{(j)} & \cdots & x_n^{(j)} \end{pmatrix}^T$. There are p nodes in the hidden layer, which share a single weight vector $\mathbf{w}$ and bias w_0. Thus the kth hidden node computes $\sigma(a_k)$ where σ is an activation function and

$$a_k = \mathbf{w}^T \mathbf{x}^{(k)} + w_0.$$

Let $\mathbf{a} = \begin{pmatrix} a_1 & \cdots & a_p \end{pmatrix}^T$. The output node then combines the hidden nodes in the usual way using weights $\mathbf{v}$ and v_0 to produce $y = \sigma(a)$ where

$$a = \sum_{i=1}^{p} v_i \sigma(a_i) + v_0.$$

Collecting all the parameters of the network into a single vector $\boldsymbol{\theta}$, the error for a single labelled example is $E(\boldsymbol{\theta})$.

(a) Show that the value of $\delta = \partial E(\boldsymbol{\theta})/\partial a$ is

$$\delta = \sigma'(a) \frac{\partial E(\boldsymbol{\theta})}{\partial y}.$$

[3 marks]

(b) Find expressions for the partial derivatives of $E(\boldsymbol{\theta})$ with respect to the parameters of the single output node. [5 marks]

(c) Show that the partial derivatives $\delta_i = \partial E(\boldsymbol{\theta})/\partial a_i$ for the hidden nodes are

$$\delta_i = \delta v_i \sigma'(a_i).$$

[5 marks]

(d) Find expressions for the partial derivatives of $E(\boldsymbol{\theta})$ with respect to the parameters of the hidden nodes. [7 marks]