

2007 Paper 4 Question 7

Artificial Intelligence I

A very simple neural network designed to solve a two-class classification problem where the classes are labelled as 0 and 1 takes input vectors $\mathbf{x}^T = (x_1 \ x_2 \ \dots \ x_n)$ and has a weight vector $\mathbf{w}^T = (w_1 \ w_2 \ \dots \ w_n)$, both with real-valued elements. It computes the function

$$f(\mathbf{w}; \mathbf{x}) = \text{sgn}(\mathbf{w}^T \mathbf{x}) = \text{sgn}\left(\sum_{i=1}^n w_i x_i\right)$$

where the function sgn is defined as

$$\text{sgn}(z) = \frac{1}{1 + e^{-z}}.$$

There exists a training sequence for the network containing m labelled examples

$$((\mathbf{x}_1, o_1), (\mathbf{x}_2, o_2), \dots, (\mathbf{x}_m, o_m))$$

where the o_i denote desired outputs and take values in $\{0, 1\}$.

- (a) For the given training sequence, the error of the network when the weights are set to \mathbf{w} is to be defined by the function

$$E(\mathbf{w}) = \lambda \|\mathbf{w}\| + \sum_{i=1}^m \left(o_i \log \frac{1}{f(\mathbf{w}; \mathbf{x}_i)} + (1 - o_i) \log \frac{1}{1 - f(\mathbf{w}; \mathbf{x}_i)} \right)$$

where λ is a fixed, real-valued parameter, we use natural logarithms, and $\|\mathbf{w}\| = \sum_{i=1}^n w_i^2$. Derive an algorithm that can be used to train this neural network by attempting to find a weight vector minimizing $E(\mathbf{w})$. [17 marks]

- (b) Describe the way in which your algorithm might be affected by applying it using different values for the parameter λ , in particular very large or very small values. [3 marks]