1998 Paper 9 Question 8

Neural Computing

When using a feed-forward network to solve a classification problem we can interpret the network's outputs as posterior probabilities of class membership, and then subsequently use these probabilities to make classification decisions. Alternatively, we can treat the network as a discriminant function which is used to make the classification decision directly. Discuss the relative merits of these two approaches. [7 marks]

Explain the concept of a likelihood function, and the principle of maximum likelihood. [3 marks]

Consider a feed-forward network which implements a function $y(\mathbf{x}, \mathbf{w})$ in which y is the output variable, $\mathbf{x}$ is the vector of input variables, and $\mathbf{w}$ is the vector of weight parameters. We wish to use this network to solve a classification problem involving two classes $\mathcal{A}$ and $\mathcal{B}$. The value of y, when the network is presented with an input vector $\mathbf{x}$, is to be interpreted as the posterior probability $P(t = 1|\mathbf{x})$ in which $t = 1$ denotes class $\mathcal{A}$ and $t = 0$ denotes class $\mathcal{B}$. Write down the probability distribution of t given y. Use the principle of maximum likelihood to derive an expression for the corresponding error function defined over a set of training data comprising input vectors $\mathbf{x_n}$ and targets t_n, where $n = 1, \ldots, N$.

Write down a suitable form for the output unit activation function $y = g(a)$. Hence evaluate the derivative of $\ln P(t|y)$ with respect to a. [10 marks]