MODEL ANSWER

Neural Computing 2002 Paper 7 Question 7 (JGD)

(Subject areas: Neural net classifiers; Content-addressable memories.)

A.

If a neural network can provide accurate assessments of posterior probabilities, then these can be used to make optimal decisions and will give the same results as the discriminant function. In addition, the interpretation as posterior probabilities allows the following advantages: (i) Corrections can be made for prior probabilities which differ between the data set used to train the network and the test data on which the trained network will be used. (E.g. in a medical screening application, one might train a network on data sets or images half of which are clinically pathological and half of which are not, whereas in real-world use the typical ratio might be more like 1-in-50.) (ii) If a non-trivial loss matrix (specifying the penalties paid for the different types of classification errors) is introduced, then optimal decisions can still be made. (iii) If the prior distributions of the test data change, or if the elements of the loss matrix are changed, then optimal decisions can still be made without the need to re-train the network. A reject option can be applied to improve the misclassification rate, at the expense of not reaching a decision for some fraction of the data.

The disadvantage of trying to estimate posterior probabilities is that typically much more training data will be required than will be needed to obtain a good discriminant function. If the supply of training data is limited, it is possible for the network to provide an accurate discriminant function, even though its estimates of posterior probabilities may be poor.

[8 marks]

B.

The likelihood function (for a parametric model that is described by a vector of parameters $\mathbf{w}$, and a given data set D) is defined by the conditional probability $p(D|\mathbf{w})$ of the data D given $\mathbf{w}$, viewed as a function of $\mathbf{w}$. The principle of maximum likelihood says that the vector of parameters $\mathbf{w}$ should be set to those values which maximize the likelihood function. These values correspond to the choice of parameters for which the observed data set is the most probable.

[2 marks]

C.

A Hopfield content-addressable, associative memory allows patterns to be stored in such a way that each one can be retrieved by using just a portion, or a corrupted version, of it as its "address." The memory functions dynamically by detecting an association between an input pattern and a stored memory, which it then retrieves in its entirety.

The network consists of N "neurones" that are fully connected to each other. Their connections have weights of adjustable strengths, defining an NxN connectivity matrix.

This matrix of weights is adjusted during a learning phase that stores the patterns in memory. Each of the N neurones has a binary state, and the N-dimensional space of their states is called the configuration space. A stored "memory" is just a particular point in this N-dimensional configuration space. The training phase consists in setting all the NxN weights such that a particular memory is a stable state consistent with all the other stable states, so that once the network reaches such a state, no neurones will change their state further. Then, whenever a new input pattern first appears (implemented by re-setting the states of the N neurones), the collective state of the network will evolve by each neurone following a threshold rule: The $N - 1$ inputs that it receives from all the other neurones (each of whose state is either a $+1$ or a -1), each multiplied by the corresponding weight for that connection, are all added together and compared against a threshold. If this inner product (sum) exceeds the threshold, then the neurone's state is set to $+1$; else it is set to -1. This process reiterates for all the N neurones until a stable state is reached, when no neurone changes its state anymore. This collective state is called an stable attractor. Each stable attractor represents one memory. The range of different input states that will all converge eventually to a particular such stable attractor is called the basin of attraction for that memory. Because the states within this basin of attraction represents similar, but corrupted (or partial) versions of the stable state, a Hopfield memory is capable of "completing" a memory from just parts of it serving as a trigger, rather like happens in human memory. Hence it is deemed content-addressable and associative. It is also capable of overcoming noise and corruption of an input pattern, provided that the input remains within the basin of attraction of the attractor. The network capacity of a Hopfield network is about $(0.15)N$, where N is the number of neurones. This capacity would be greater if only orthogonal patterns were stored.

[10 marks]