1999

p9q9

JGD

Model Answer, Neural Computing, Question 3

(a) *The curse of dimensionality.*

Many simple models used for pattern recognition have the unfortunate property that the number of adaptive parameters in the model grows rapidly, sometimes exponentially, with the number of input variables (i.e. with the dimensionality of the input space). Since the size of the data set must grow with the number of parameters, this leads to the requirement for excessively large data sets, as well as increasing the demands on computational resources. An important class of such models is based on linear combinations of fixed, non-linear basis functions. The worst aspects of the curse of dimensionality in such models can be alleviated, at the expense of greater computational complexity, by allowing the basis functions themselves to be adaptive.

(b) *The Perceptron.*

The Perceptron is a simple neural network model developed in the 1960s by Rosenblatt. He built hardware implementations of the Perceptron, and also proved that the learning algorithm is guaranteed to find an exact solution in a finite number of steps, provided that such a solution exists. The limitations of the Perceptron were studied mathematically by Minsky and Papert.

(c) *Error back-propagation.*

Neural networks consisting of more than one layer of adaptive connections can be trained by error function minimisation using gradient-based optimisation techniques. In order to apply such techniques, it is necessary to evaluate the gradient of the error function with respect to the adaptive parameters in the network. This can be achieved using the chain rule of calculus which leads to the error back-propagation algorithm. The name arises from the graphical interpretation of the algorithm in terms of a backwards propagation of error signals through the network.

(d) *Generalisation.*

The parameters of a neural network model can be determined through minimisation of an appropriate error function. However, the goal of training is not to give good performance on the training data, but instead to give the best performance (in terms of smallest error) on independent, unseen data drawn from the same distribution as the training data. The capability of the model to give good results on unseen data is termed generalisation.

(e) *Loss matrix.*

In many classification problems, different misclassification errors can have different consequences and hence should be assigned different penalties. For example, in a medical screening application the cost of predicting that a patient is normal when in fact they have cancer is much more serious than predicting they have cancer when in fact they are healthy. This effect can be quantified using a loss matrix consisting of penalty values for each possible combination of true class versus predicted class. The elements on the leading diagonal correspond to correct decisions and are usually chosen to be zero. A neural network model can be used to estimate the posterior probabilities of class membership for a given input vector. Simple decision theory then shows that if these posterior probabilities are weighted by the appropriate elements of the loss matrix, then selection of the largest weighted probability represents the optimal classification strategy in the sense of minimising the average loss.