# Literature Review

Handwriting recognition has been the major subject of research for almost the last forty years. The topic has seen major development in recent years.

**// Pattern Recognition(Bishop)**

In Chapters 3 and 4 we considered models for regression and classification that comprised linear combinations of fixed basis functions. We saw that such models have

useful analytical and computational properties but that their practical applicability

was limited by the curse of dimensionality. In order to apply such models to largescale problems, it is necessary to adapt the basis functions to the data.

Support vector machines (SVMs), discussed in Chapter 7, address this by first

defining basis functions that are centred on the training data points and then selecting

a subset of these during training. One advantage of SVMs is that, although the

training involves nonlinear optimization, the objective function is convex, and so the

solution of the optimization problem is relatively straightforward. The number of

basis functions in the resulting models is generally much smaller than the number of

training points, although it is often still relatively large and typically increases with

the size of the training set. The relevance vector machine, discussed in Section 7.2,

also chooses a subset from a fixed set of basis functions and typically results in much

sparser models. Unlike the SVM it also produces probabilistic outputs, although this

is at the expense of a nonconvex optimization during training.

<Pattern Recognition: Introduction of NN>

An alternative approach is to fix the number of basis functions in advance but

allow them to be adaptive, in other words to use parametric forms for the basis functions in which the parameter values are adapted during training. The most successful

model of this type in the context of pattern recognition is the feed-forward neural

network, also known as the multilayer perceptron, discussed in this chapter. In fact,

‘multilayer perceptron’ is really a misnomer, because the model comprises multiple layers of logistic regression models (with continuous nonlinearities) rather than

multiple perceptrons (with discontinuous nonlinearities). For many applications, the

resulting model can be significantly more compact, and hence faster to evaluate, than

a support vector machine having the same generalization performance. The price to

be paid for this compactness, as with the relevance vector machine, is that the likelihood function, which forms the basis for network training, is no longer a convex

function of the model parameters. In practice, however, it is often worth investing

substantial computational resources during the training phase in order to obtain a

compact model that is fast at processing new data.

**The term ‘neural network’ has its origins in attempts to find mathematical representations of information processing in biological systems (McCulloch and Pitts,**

**1943; Widrow and Hoff, 1960; Rosenblatt, 1962; Rumelhart et al., 1986)**

**// Use of backpropagation in NN**

The paper "Learning internal representations by error propagation" by David Rumelhart, Geoffrey Hinton, and Ronald Williams, published in 1986, suggested that neural networks with backpropagation could be used for character recognition. This paper introduced the backpropagation algorithm, which is a method for training neural networks to minimize error on a supervised learning task, such as character recognition. Prior to this paper, neural networks had been used for character recognition, but the backpropagation algorithm made it possible to train large and deep networks more effectively, which led to significant improvements in performance.

**// State of the art in Low Resource language**

**// In Non-Latin Language**