mean—a reasonable working assumption. In contrast, an MLP becomes more certain in its response when far-flung data are used. Whether this is an advantage or disadvantage depends largely on the application, but on the whole the MLP’s uncritical extrapolation is regarded as a bad point; extrapolation far from training data is usually dangerous and unjustified. However, both methods, like logistic regression, are far better at extrapolation than methods like regression or polynomial networks that have no constraints on the output estimate.

RBFs are also more sensitive to the curse of dimensionality and have greater difficulties if the number of input units is large.

Automated Neural Nets

Several data mining tools offer neural nets that have “smart” search algorithms to choose the appropriate starting points for their parameters. But the biggest benefit of these algorithms is that they search over the decision surface with different initial learning rates (which also decay between iterations), different momentums, and different number of nodes in the middle layer. Usually, you have to choose the number of middle layers to use before the algorithm takes over. Both SPSS Clementine and STATISTICA Data Miner have very powerful automated neural nets.

GENERALIZED ADDITIVE MODELS (GAMs)

As theory of Generalized Linear Models (GLMs) developed in the 1980s, the need for an increasing number of predictor variables was recognized as a key issue. The problem with increasing the number of predictor variables is that the variance increases also. The higher the variance, the harder it is for a prediction algorithm to perform well (perform acceptably on new data). This is one aspect of the “curse of dimensionality.” To bypass this problem, Stone (1986) proposed modification of the GLM by replacing the definition of each predictor variable with an additive approximation term. This approximation is performed with a linear univariate smoothing function. This approach avoided the curse of dimensionality by performing a simple fitting of each predictor variable to the dependent variable. The new approach also expressed the definition of each predictor variable such that it was possible to relate how the variable affected the dependent variable. Remember, in the standard Multiple Linear Regression (MLR) equation, the estimated coefficients represent effects of differing scale, as well as differing relationships to the dependent variable. Consequently, you can’t analyze the MLR coefficients directly to determine relationships. But with the enhancement by Stone, you can see these relationships directly. Still, the cost of that enhancement was a decrease in generalization (the ability to perform acceptably on new data).

Hastie and Tibshirani (1990) incorporated Stone’s idea into a formal definition of Generalized Additive Models (GAMs). A GAM uses a nonlinear link function to map input data into a solution space, similar to a GLM. This flexible approach to mapping of inputs can fit the response probability distribution of any member of the exponential family of