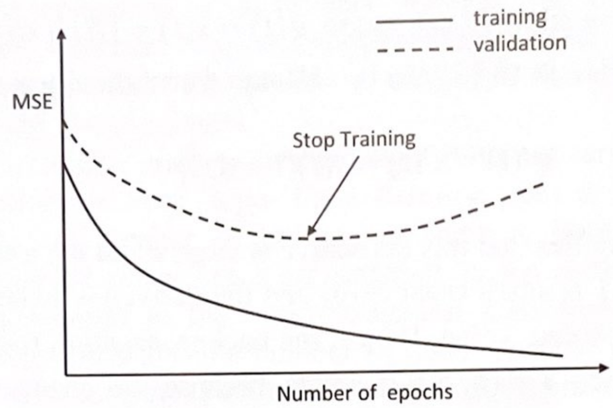

FIGURE 6.6 Early stopping technique for model selection in MLP networks.

choose (vary) just 1 or 2 parameters for model selection, while setting all other factors to some (reasonable) fixed values. Typical software packages include about half-a-dozen 'user-defined' parameters that affect model complexity. So practitioners need to have some knowledge and understanding of the MLP software they use for modeling.

Practical model selection strategies involve selecting the number of hidden units and using the 'early stopping' rules. The latter approach, which is specific to MLP networks using backpropagation, is discussed next. The early stopping technique controls the model complexity as follows: a large over-parameterized MLP network is trained for a while, and its generalization performance is evaluated using a separate validation set. This validation set is used to determine when to stop training. That is, training should be stopped when the validation error reaches its minimum, as in Fig. 6.6. In practice, using a separate validation set may be wasteful, so the validation error is estimated via resampling.

***Example 6.1**: Application of MLP network for univariate regression.*
This example illustrates MLP modeling of a synthetic data set where:

- the input values x are uniformly distributed in the range [0, 1],
- the output values are generated as $y = \sin^2(2\pi x) + \xi$,
- the additive noise ξ is Gaussian with standard deviation 0.1.

30 samples are used for training MLP network. The MLP software used in this experiment implements standard backpropagation algorithm

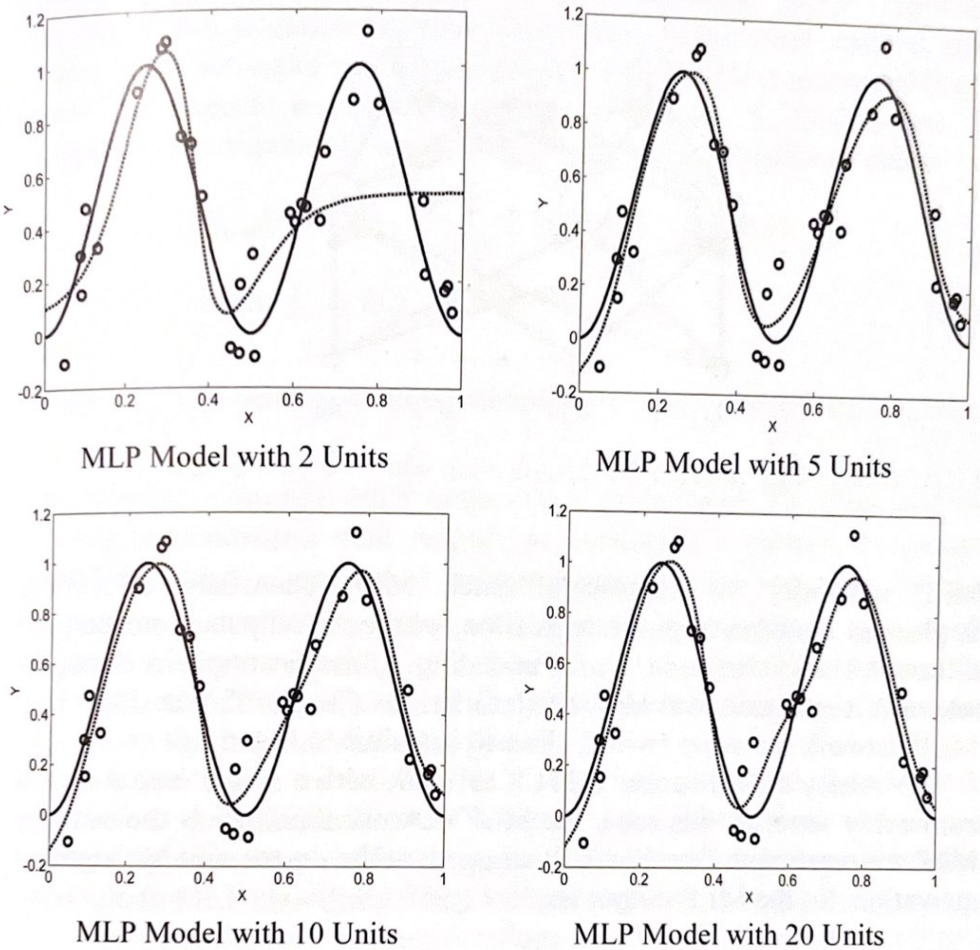

FIGURE 6.7 MLP network for univariate regression. The true target function is shown in solid line and its estimate in dashed line.

trying to achieve thorough minimization of MSE for training data. The number of hidden units m is the only user-defined complexity parameter. Several MLP models for $m = 2$, 5, 10 and 20 are shown in Fig. 6.7. From this figure, the model with 2 hidden units underfits the data. MLP models with 5 and 10 units closely approximate the target function; and the model with 20 units shows slight overfitting for input values near 0.5. There is almost no overfitting, even when the number of hidden units is quite large.

MLP networks can be also used for classification. MLP classifiers often use the squared loss, so they are computationally equivalent to