with these algorithms during the 90-day license period, even if you normally use a different package in your organization.

ADVANCED DATA MINING ALGORITHMS

Interactive Trees

The STATISTICA Interactive Trees (I-Trees) module builds trees for predicting dependent variables that are continuous (estimation) or categorical (classification). The program supports the classic Classification and Regression Tree (CART) algorithm (Breiman et al.; see also Ripley, 1996) as well as the CHAID algorithm (Chi-Square Automatic Interaction Detector; Kass, 1980). The module can use algorithms, user-defined rules, criteria specified via an interactive graphical user interface (brushing tools), or a combination of those methods. This enables users to try various predictors and splitting criteria in combination with almost all the functions of automatic tree building.

Figure 8.2 displays the tree results layout in the I-Trees module. The main results screen is shown in Figure 8.3.

Manually Building the Tree

The I-Trees module doesn’t build trees by default, so when you first display the Trees Results dialog, no trees have been built. (If you click the Tree Graph button at this point, a single box will be displayed with a single root node, as in Figure 8.4.)

The Tree Browser

The final tree results are displayed in the workbook tree browser, which clearly identifies the number of splitting nodes and terminal nodes of the tree (Figure 8.5).

To review the statistics and other information (e.g., splitting rule) associated with each node, simply highlight it and review the summary graph in the right pane. The split nodes can be collapsed or expanded in the manner that most users are accustomed to from standard MS Windows-style tree browser controls. Another useful feature of the workbook tree browser is the ability to quickly review the effect of consecutive splits on the resulting child nodes in an animation-like manner.

Advantages of I-Trees

- I-Trees is particularly optimized for very large data sets, and in many cases the raw data do not have to be stored locally for the analyses.
- It is more flexible in the handling of missing data. Because the Interactive Trees module does not support ANCOVA-like design matrices, it is more flexible in the handling of missing data; for example, in CHAID analyses, the program will handle predictors one at a time to determine a best (next) split; in the General CHAID (GCHAID) Models module, observations with missing data for any categorical predictor are eliminated from the analysis.
- You can perform “what-if” analyses to gain better insights into your data by interactively deleting individual branches, growing other branches, and observing various results statistics for the different trees (tree models).