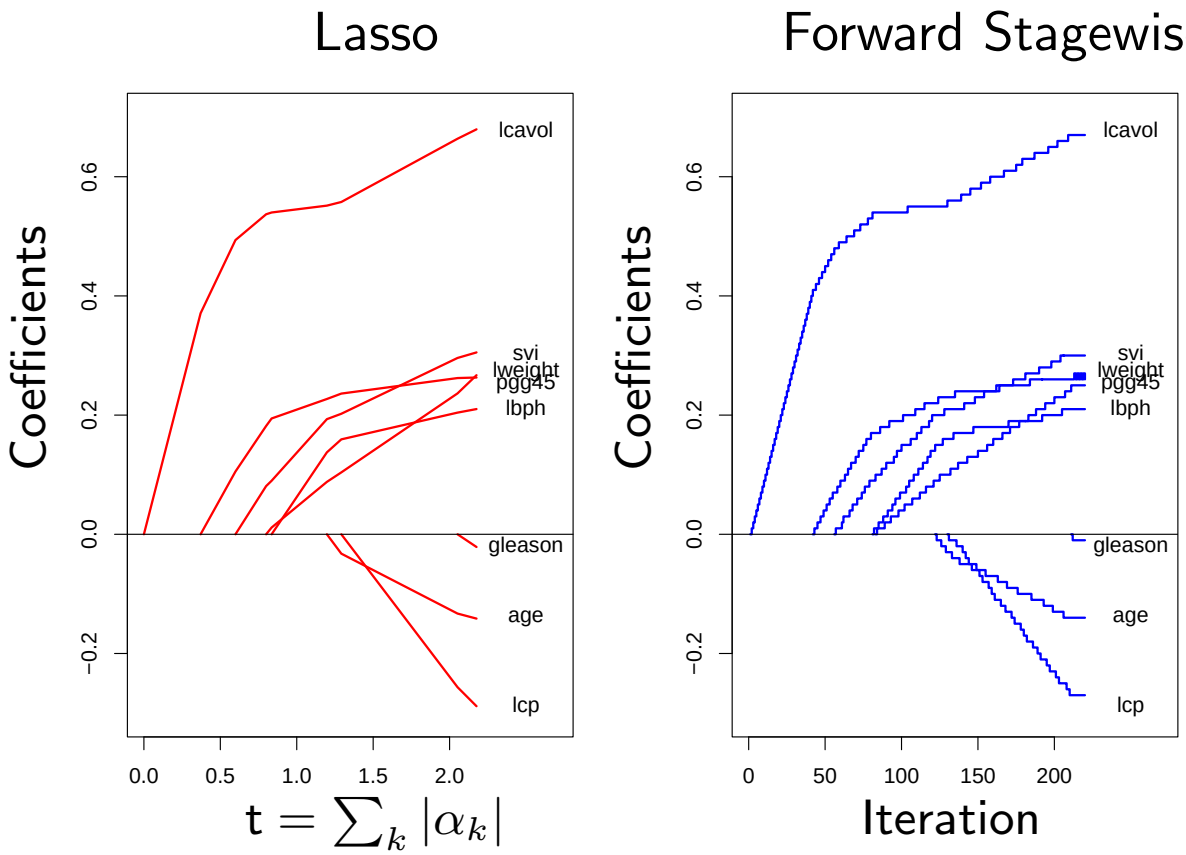

FIGURE 16.1. *Profiles of estimated coefficients from linear regression, for the prostate data studied in Chapter 3. The left panel shows the results from the lasso, for different values of the bound parameter* $t = \sum_k |\alpha_k|$. *The right panel shows the results of the stagewise linear regression Algorithm 15.1, using* $M = 220$ *consecutive steps of size* $\varepsilon = .01$.

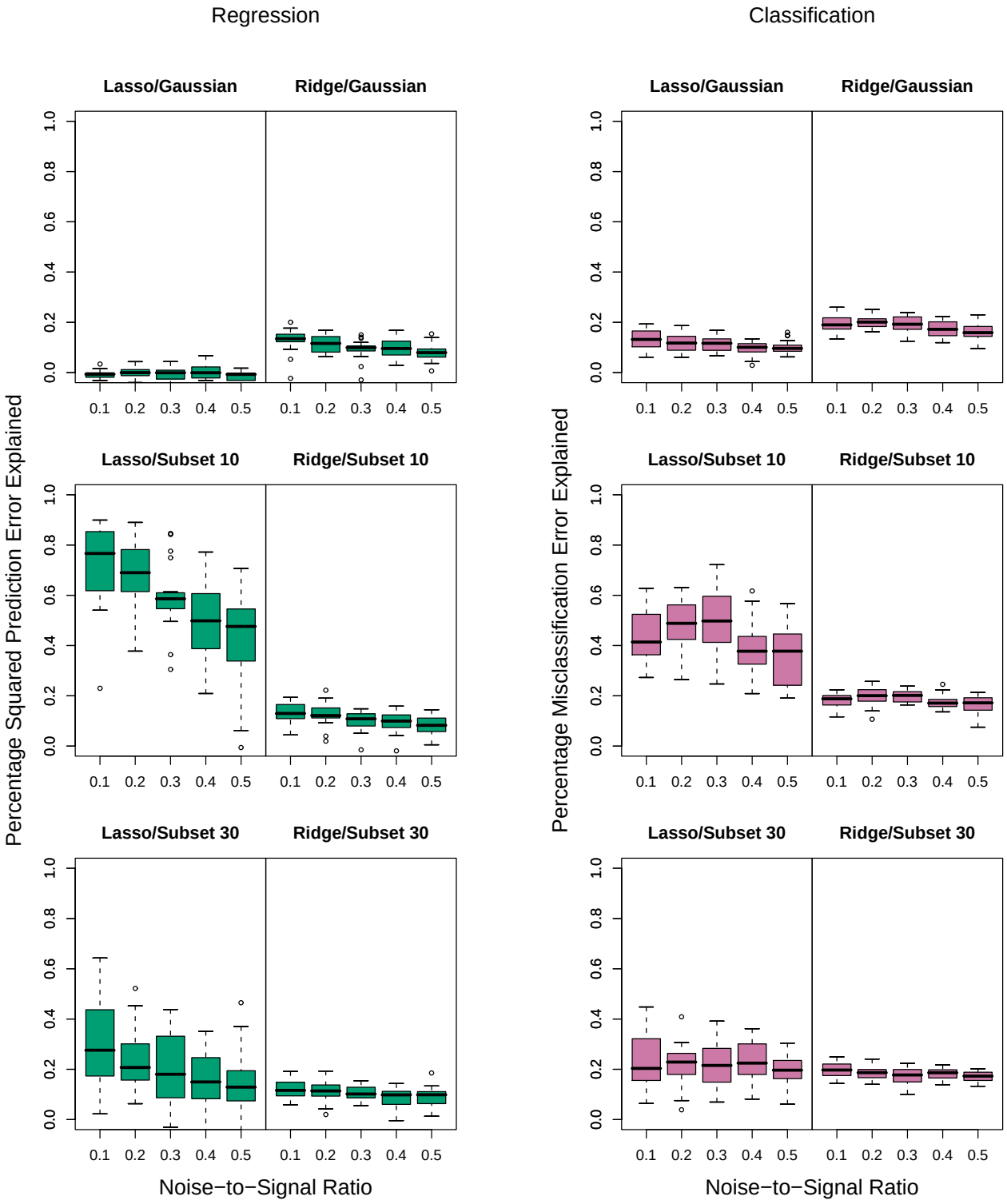

FIGURE 16.2. *Simulations that show the superiority of the L_1 (lasso) penalty over L_2 (ridge) in regression and classification. Each run has 50 observations with 300 independent Gaussian predictors. In the top*

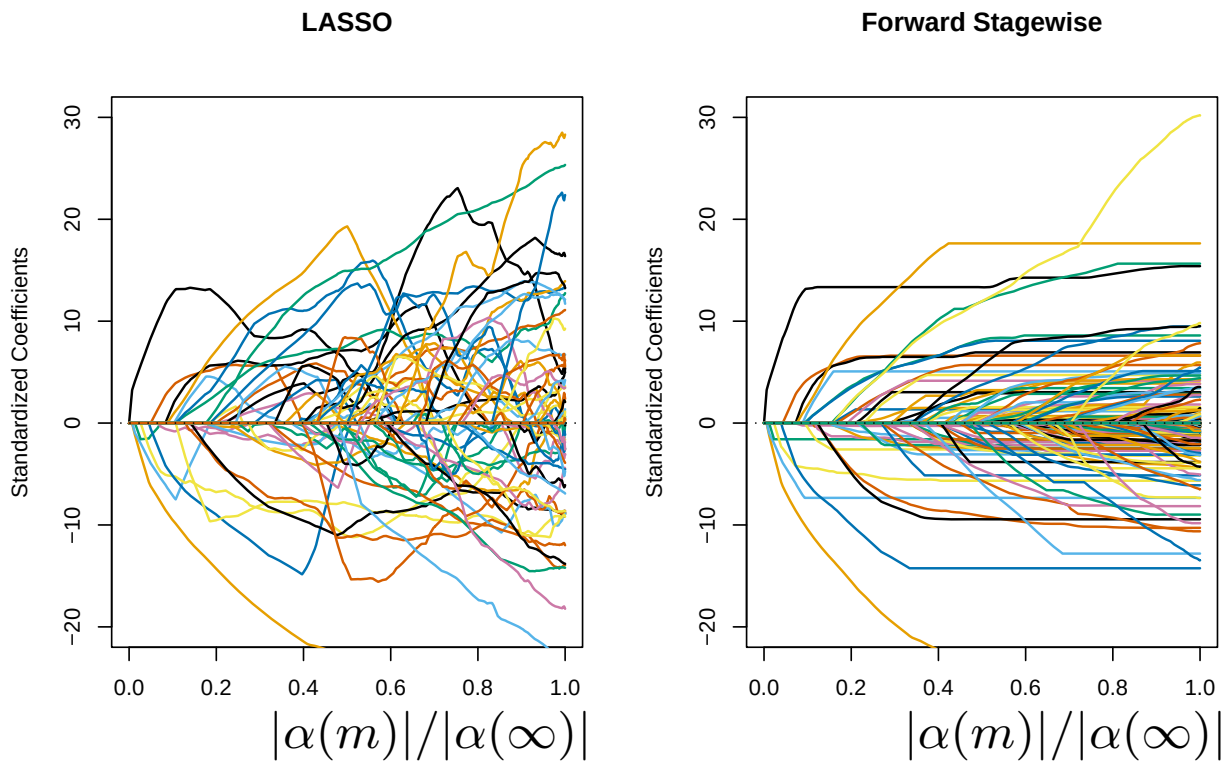

FIGURE 16.3. *Comparison of lasso and infinitesimal forward stagewise paths on simulated regression data. The number of samples is* 60 *and the number of variables is* 1000. *The forward-stagewise paths fluctuate less than those of lasso in the final stages of the algorithms.*

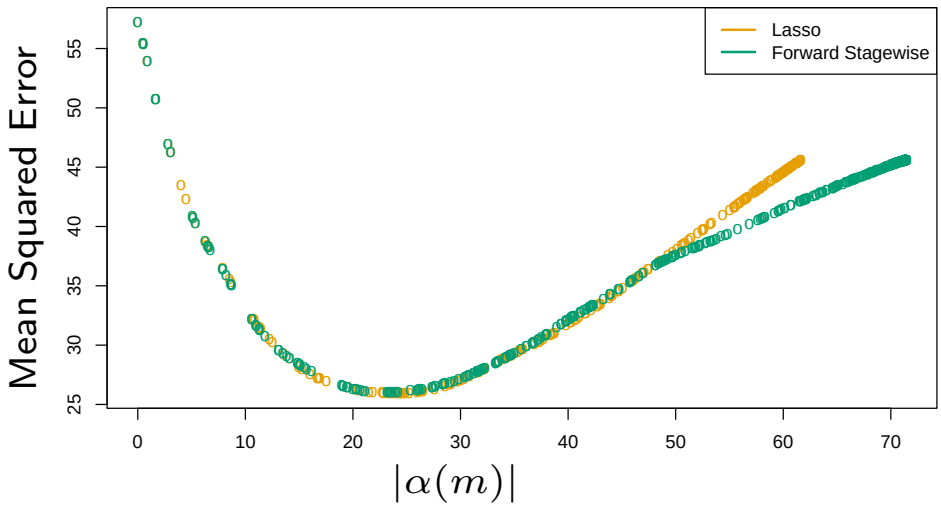

FIGURE 16.4. *Mean squared error for lasso and infinitesimal forward stagewise on the simulated data. Despite the difference in the coefficient paths, the two models perform similarly over the critical part of the regularization path. In the right tail, lasso appears to overfit more rapidly.*

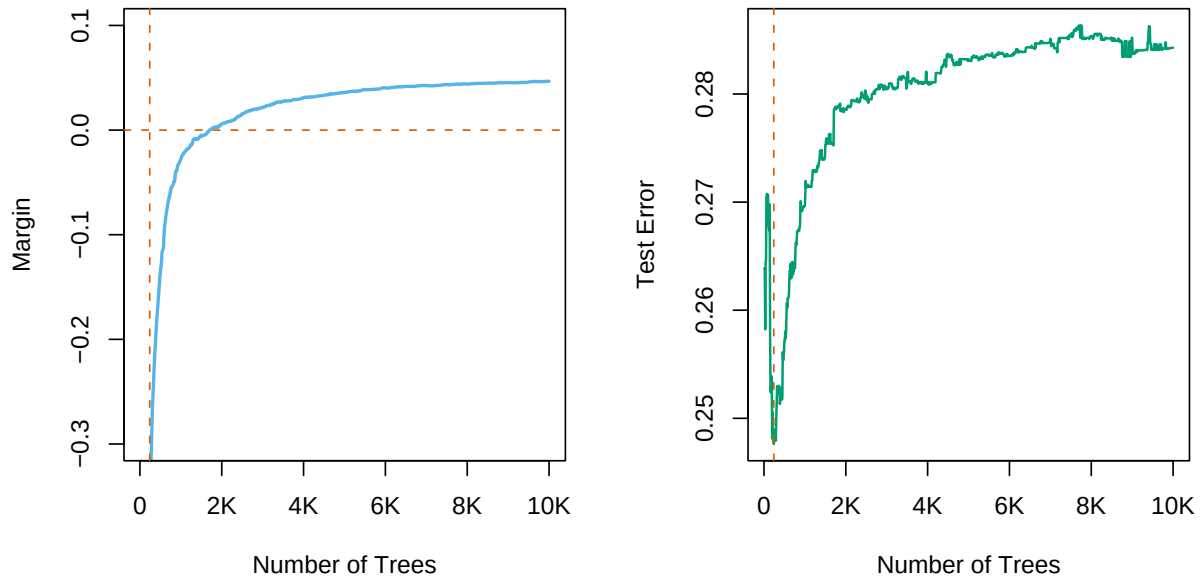

FIGURE 16.5. *The left panel shows the L_1 margin $m(f)$ for the Adaboost classifier on the mixture data, as a function of the number of 4-node trees. The model was fit using the R package* `gbm`*, with a shrinkage factor of* 0.02*. After* $10,000$ *trees, $m(f)$ has settled down. Note that when the margin crosses zero, the training error becomes zero. The right panel shows the test error, which is minimized at* 240 *trees. In this case, Adaboost overfits dramatically if run to convergence.*

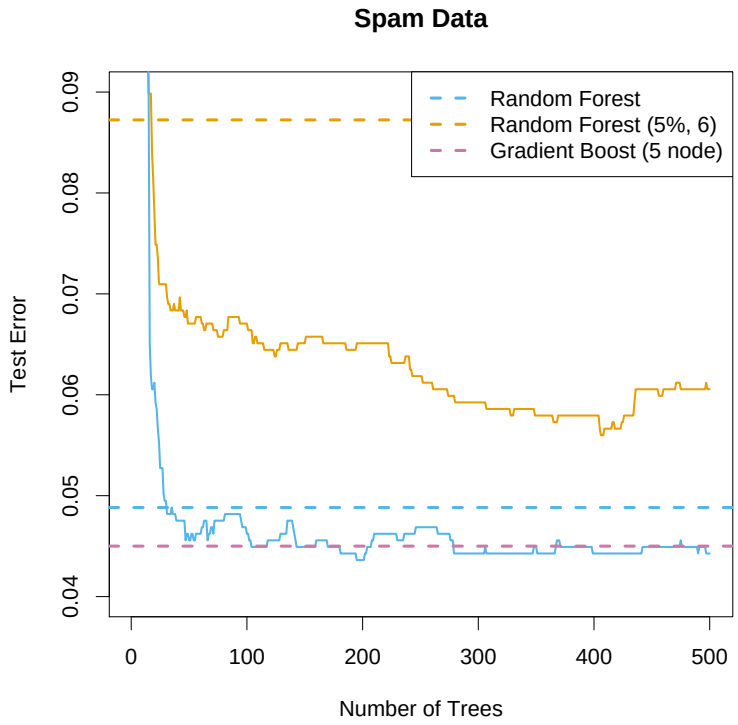

FIGURE 16.6. *Application of the lasso post-processing (15.9) to the spam data. The horizontal blue line is the test error of a random forest fit to the spam data, using* 1000 *trees grown to maximum depth (with* $m = 7$*; see Algorithm 17.1). The jagged blue curve is the test error after post-processing the first* 500 *trees using the lasso, as a function of the number of trees with nonzero coefficients. The orange curve/line use a modified form of random forest, where a random draw of* 5% *of the data are used to grow each tree, and the trees are forced to be shallow (typically six terminal nodes). Here the post-processing offers much greater improvement over the random forest that generated the ensemble.*

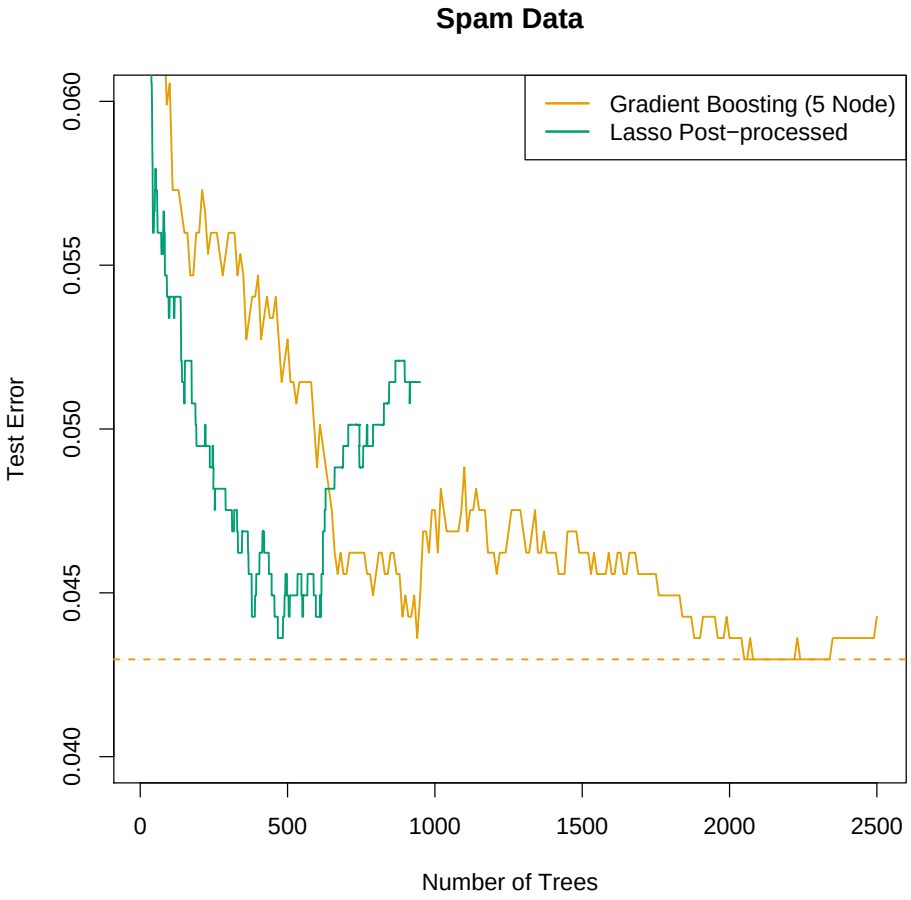

FIGURE 16.7. *Importance sampling learning ensemble (ISLE) fit to the spam data. Here we used* $\eta = 1/2$, $\nu = 0.05$, *and trees with five terminal nodes. The lasso post-processed ensemble does not improve the prediction error in this case, but it reduces the number of trees by a factor of five.*

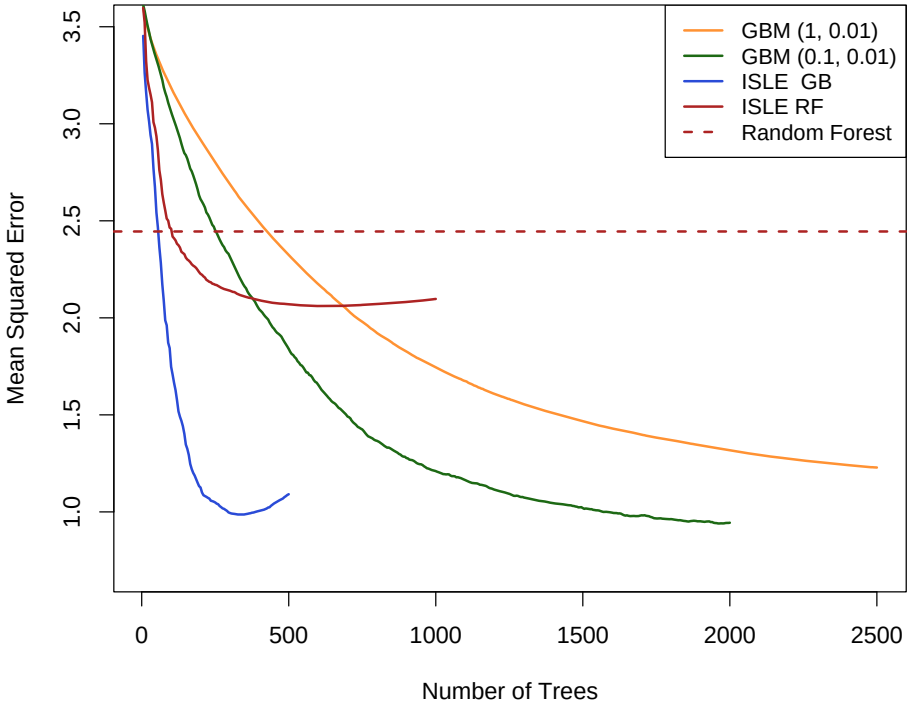

FIGURE 16.8. *Demonstration of ensemble methods on a regression simulation example. The notation GBM (0.1, 0.01) refers to a gradient boosted model, with parameters* (η, ν)*. We report mean-squared error from the true (known) function. Note that the sub-sampled GBM model (green) outperforms the full GBM model (orange). The lasso post-processed version achieves similar error. The random forest is outperformed by its post-processed version, but both fall short of the other models.*

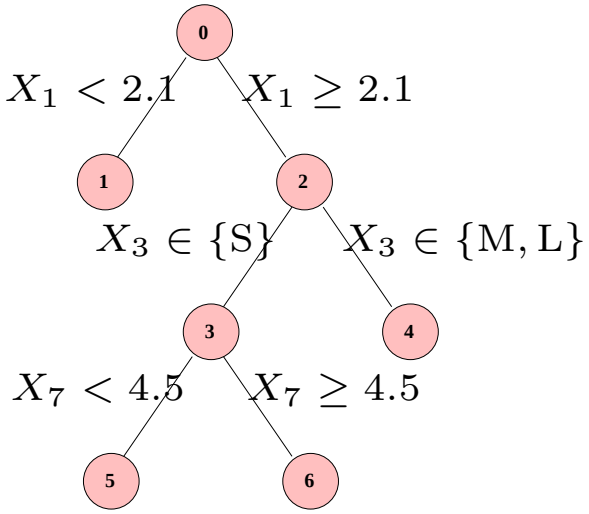

FIGURE 16.9. *A typical tree in an ensemble, from which rules can be derived.*

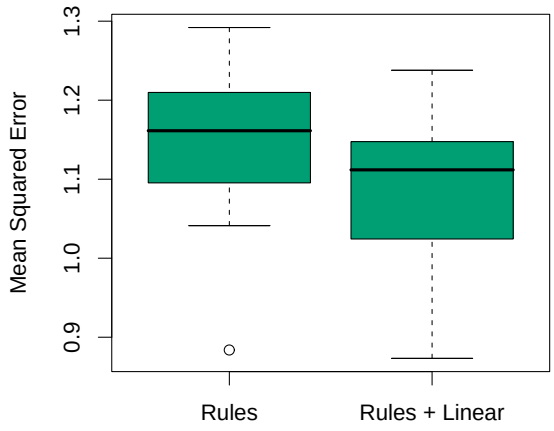

FIGURE 16.10. *Mean squared error for rule ensembles, using* 20 *realizations of the simulation example (15.13).*