

FIGURE 9.1. Spam analysis: estimated functions for significant predictors. The rug plot along the bottom of each frame indicates the observed values of the corresponding predictor. For many of the predictors the nonlinearity picks up the discontinuity at zero.

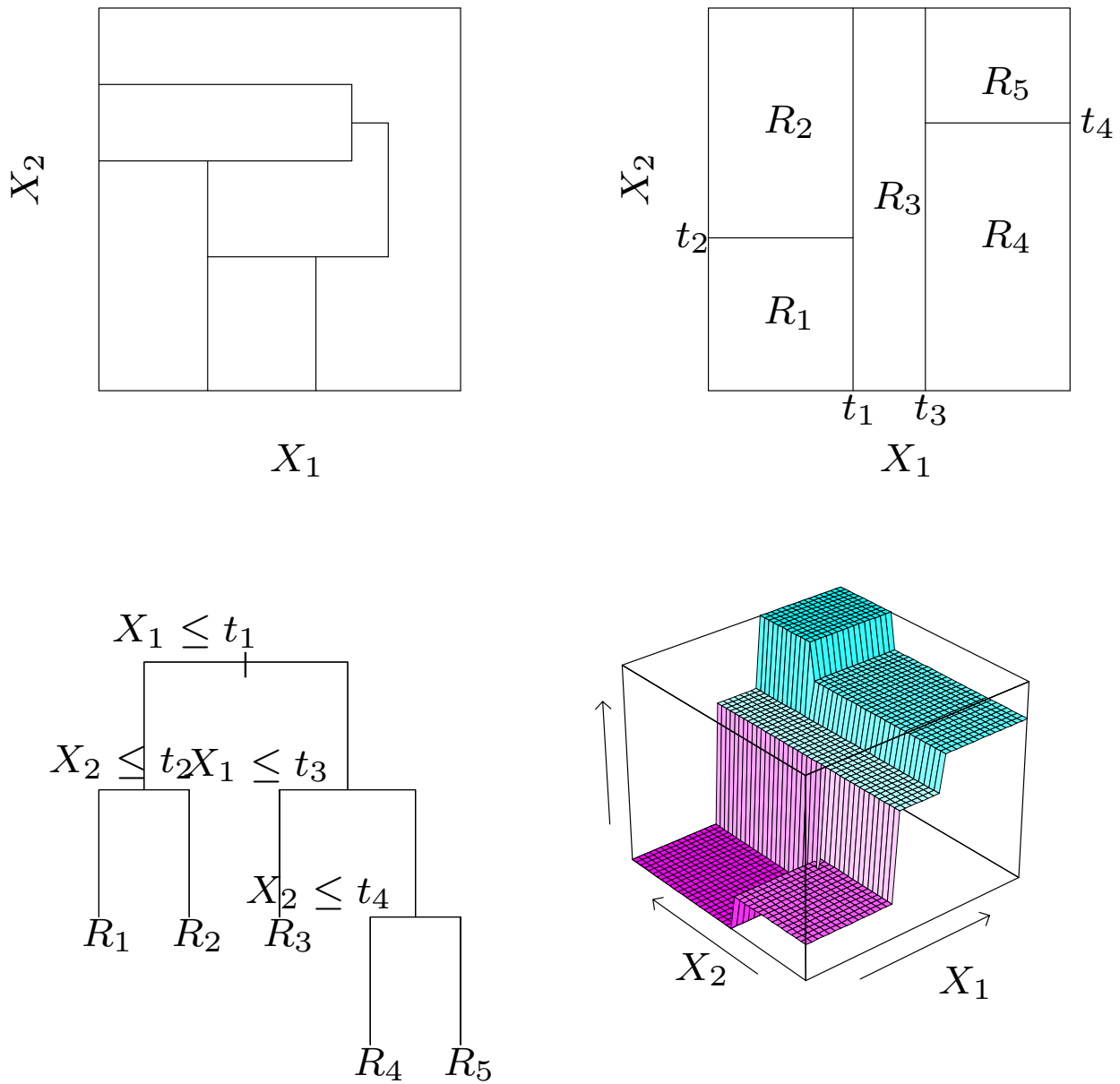


FIGURE 9.2. *Partitions and CART.* Top right panel shows a partition of a two-dimensional feature space by recursive binary splitting, as used in CART, applied to some fake data. Top left panel shows a general partition that cannot be obtained from recursive binary splitting. Bottom left panel shows the tree corresponding to the partition in the top right panel, and a perspective plot of the prediction surface appears in the bottom right panel.

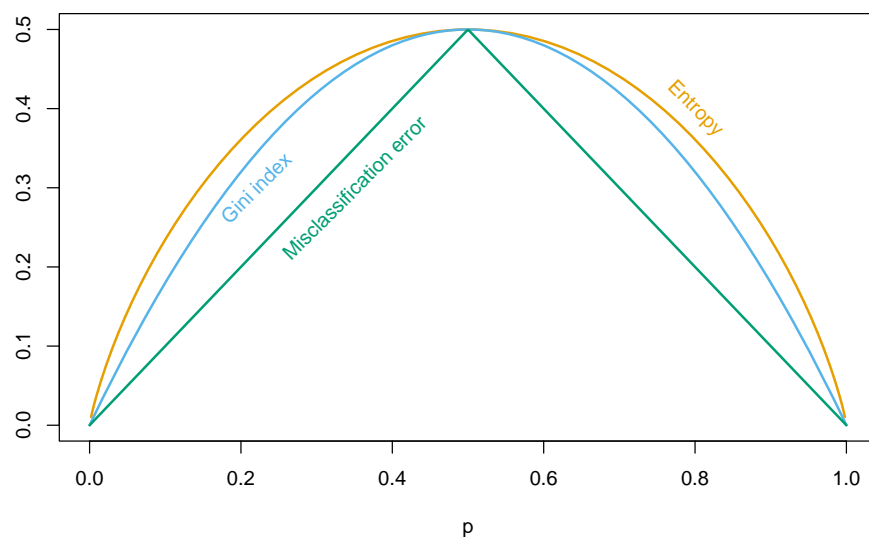


FIGURE 9.3. Node impurity measures for two-class classification, as a function of the proportion p in class 2. Cross-entropy has been scaled to pass through $(0.5, 0.5)$.

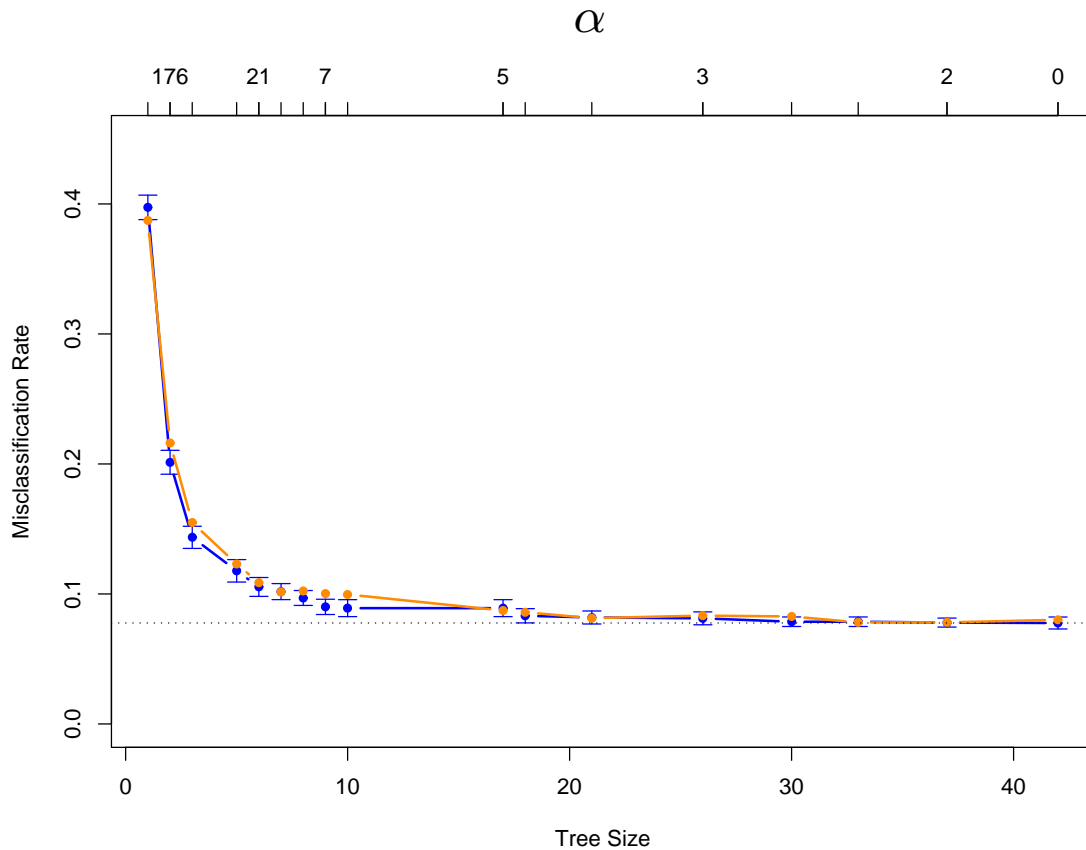


FIGURE 9.4. Results for spam example. The blue curve is the 10-fold cross-validation estimate of misclassification rate as a function of tree size, with standard error bars. The minimum occurs at a tree size with about 17 terminal nodes (using the “one-standard-error” rule). The orange curve is the test error, which tracks the CV error quite closely. The cross-validation is indexed by values of α , shown above. The tree sizes shown below refer to $|T_\alpha|$, the size of the original tree indexed by α .

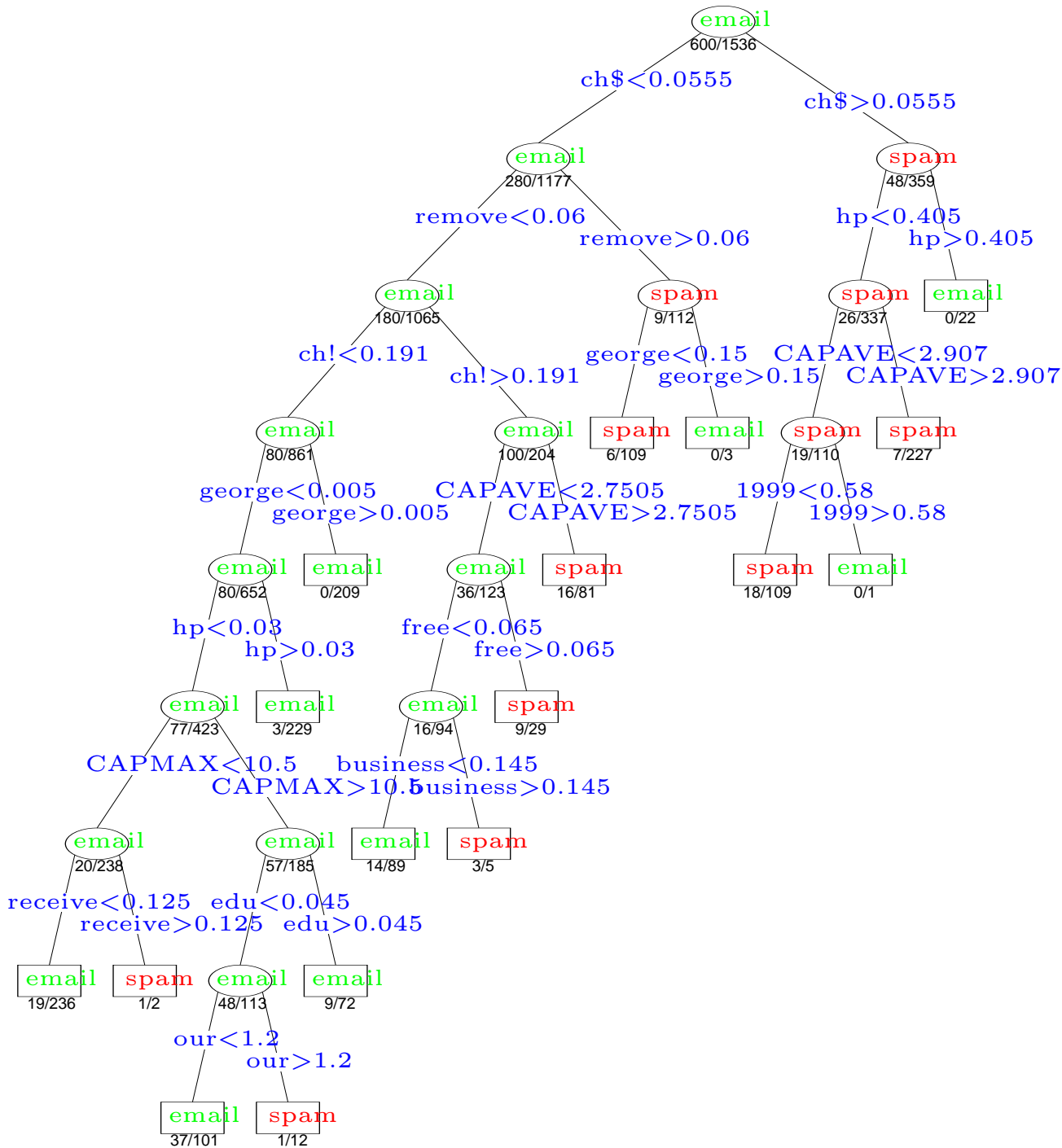


FIGURE 9.5. *The pruned tree for the spam example. The split variables are shown in blue on the branches, and the classification is shown in every node. The numbers under the terminal nodes indicate misclassification rates on the test data.*

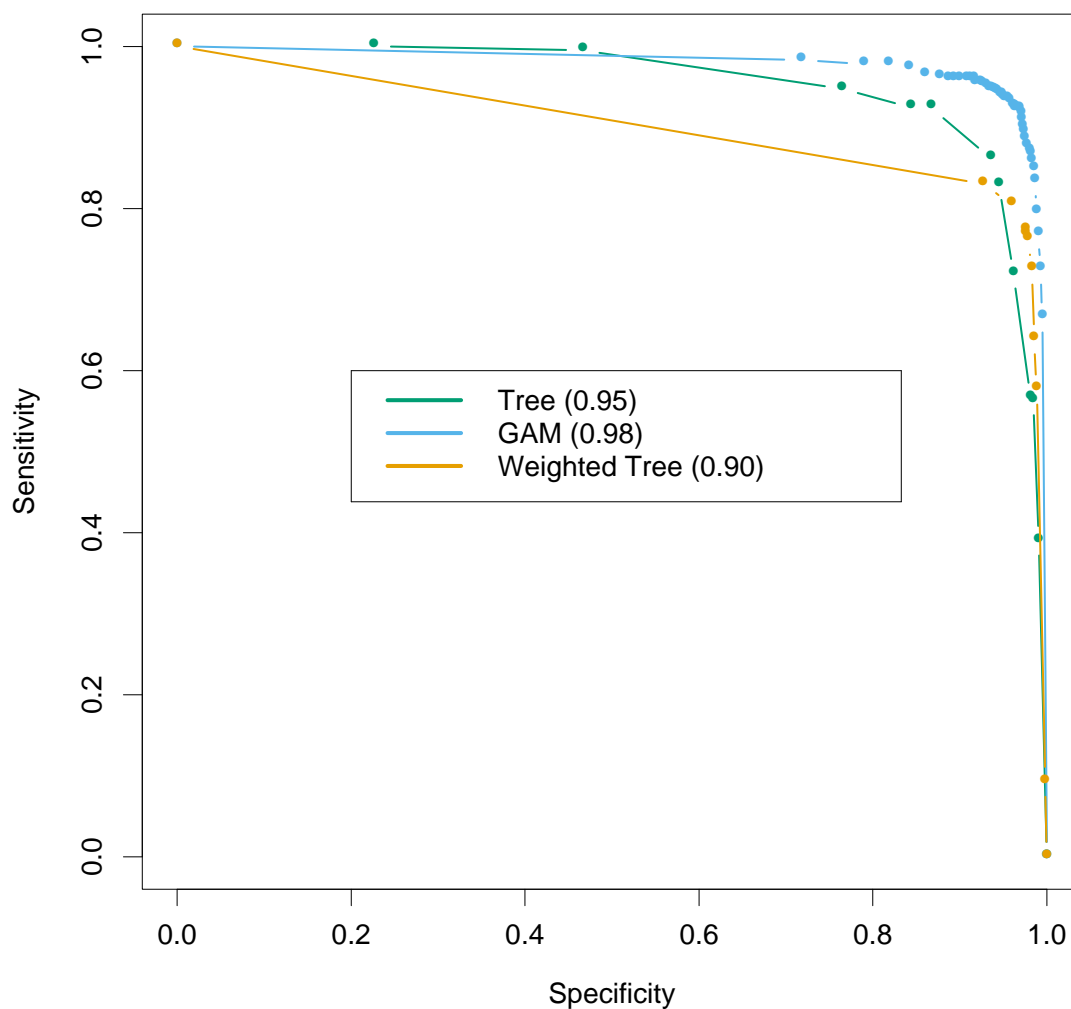


FIGURE 9.6. *ROC curves for the classification rules fit to the spam data. Curves that are closer to the north-east corner represent better classifiers. In this case the GAM classifier dominates the trees. The weighted tree achieves better sensitivity for higher specificity than the unweighted tree. The numbers in the legend represent the area under the curve.*

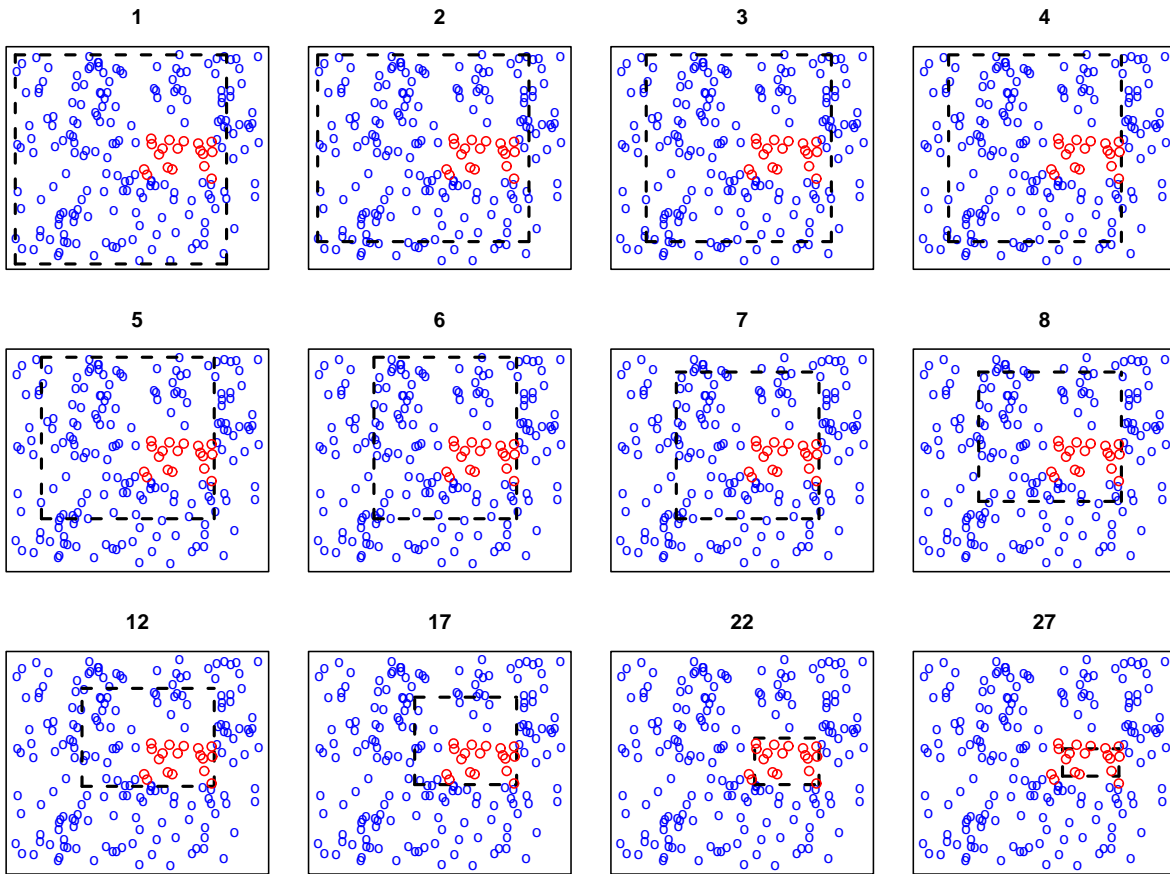


FIGURE 9.7. *Illustration of PRIM algorithm. There are two classes, indicated by the blue (class 0) and red (class 1) points. The procedure starts with a rectangle (broken black lines) surrounding all of the data, and then peels away points along one edge by a prespecified amount in order to maximize the mean of the points remaining in the box. Starting at the top left panel, the sequence of peelings is shown, until a pure red region is isolated in the bottom right panel. The iteration number is indicated at the top of each panel.*

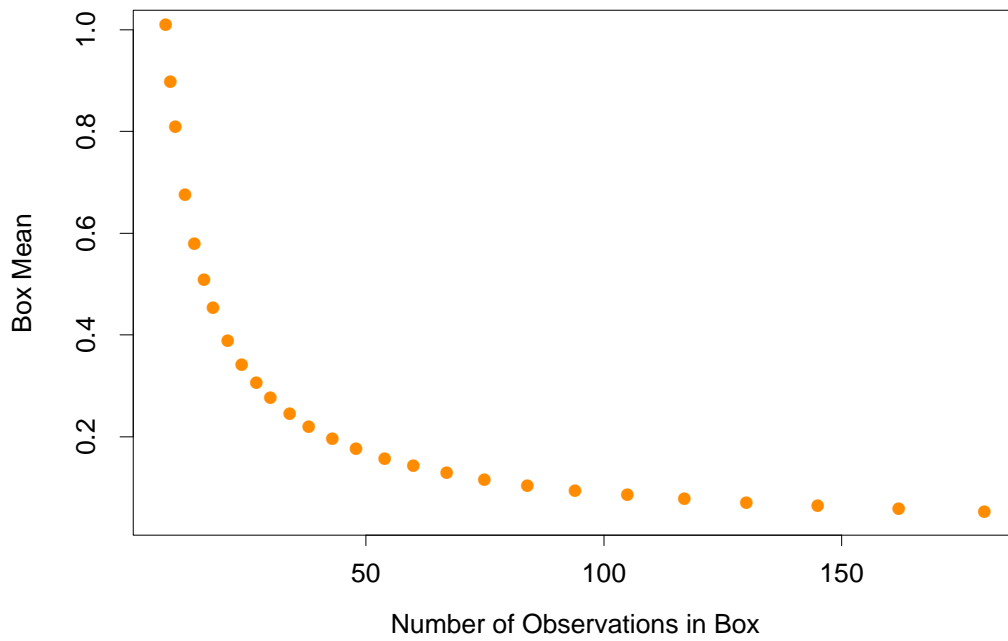


FIGURE 9.8. *Box mean as a function of number of observations in the box.*

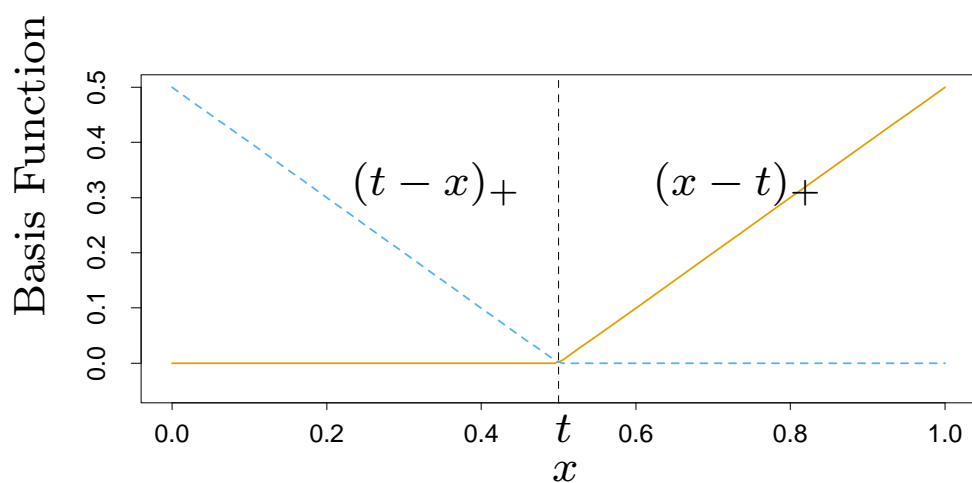


FIGURE 9.9. *The basis functions $(x - t)_+$ (solid orange) and $(t - x)_+$ (broken blue) used by MARS.*

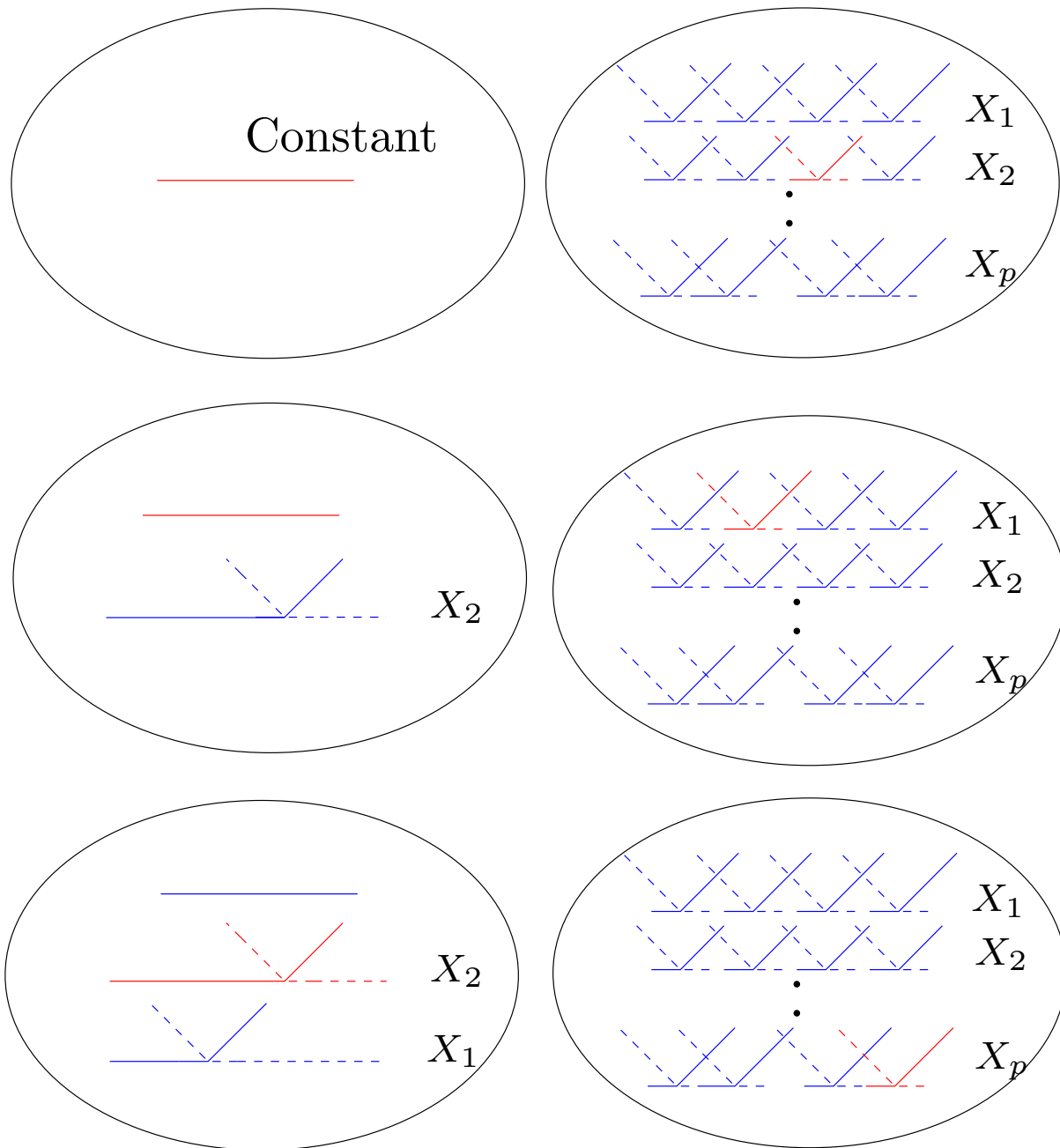


FIGURE 9.10. Schematic of the MARS forward model-building procedure. On the left are the basis functions currently in the model: initially, this is the constant function $h(X) = 1$. On the right are all candidate basis functions to be considered in building the model. These are pairs of piecewise linear basis functions as in Figure 9.9, with knots t at all unique observed values x_{ij} of each predictor X_j . At each stage we consider all products of a candidate pair with a basis function

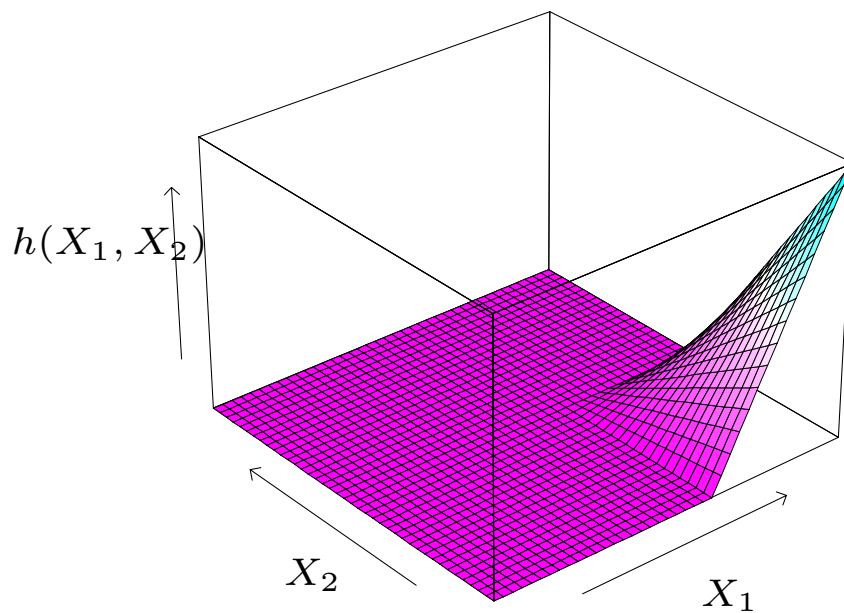


FIGURE 9.11. *The function $h(X_1, X_2) = (X_1 - x_{51})_+ \cdot (x_{72} - X_2)_+$, resulting from multiplication of two piecewise linear MARS basis functions.*

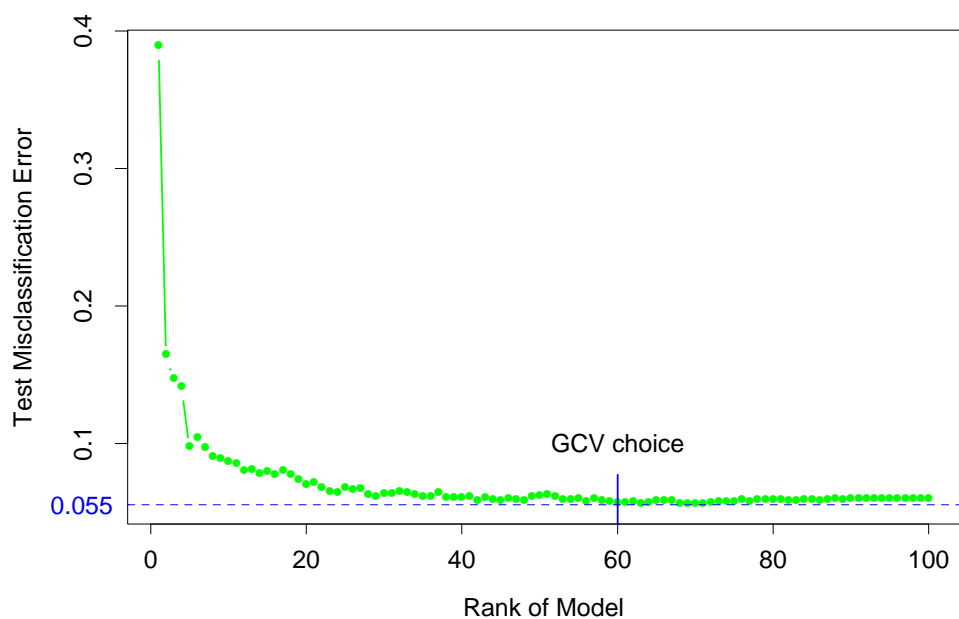


FIGURE 9.12. *Spam data: test error misclassification rate for the MARS procedure, as a function of the rank (number of independent basis functions) in the model.*

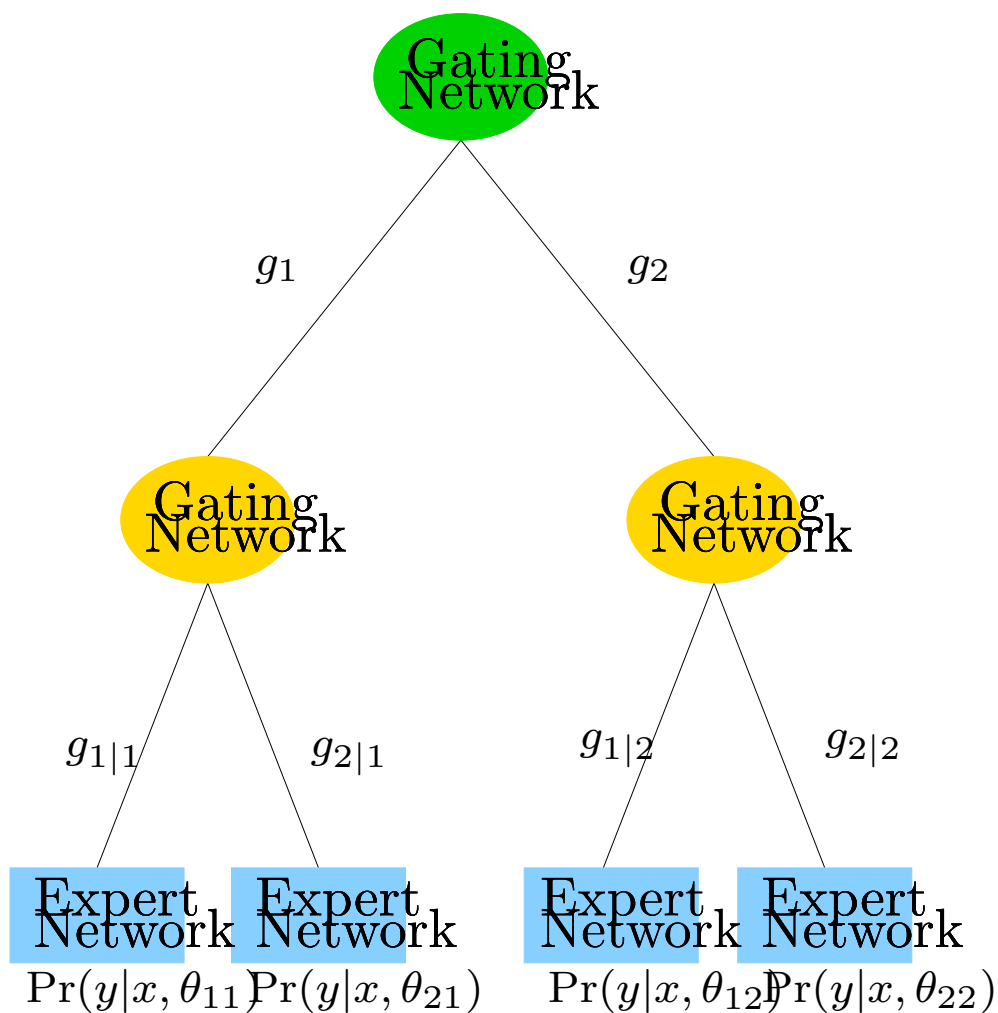


FIGURE 9.13. A two-level hierarchical mixture of experts (HME) model.