“Variable Importance” for Tree Predictors

Stephen Vardeman
Analytics Iowa LLC
ISU Statistics and IMSE

Loss for a rectangle

Consider assigning measures of "importance" of input variables for a tree predictor. In the spirit of linear models assessment of the importance of a predictor in terms of some reduction it provides in some error sum of squares, Breiman suggested the following.

Suppose that in a regression or classification tree, input variable x_j provides the rectangle splitting criterion for nodes $node_{1j}, \ldots, node_{m(j)j}$

and that before splitting at $node_{lj}$, the relevant rectangle R_{lj} has (for $\hat{y}_{lj}$ the prediction fit for that rectangle) associated sum of training losses

$$E_{lj} = \sum_{i \text{ with } \mathbf{x}_i \in R_{lj}} L\left(\hat{y}_{lj}, y_i\right)$$

Loss reduction from splitting a rectangle

Then suppose that after splitting R_{lj} on variable x_j to create rectangles R_{lj}^1 and R_{lj}^2 (with respective fitted predictions $\hat{y}_{lj}^1$ and $\hat{y}_{lj}^2$) one has sums of training losses associated with those two rectangles

$$E_{lj}^1 = \sum_{i \text{ with } \mathbf{x}_i \in R_{lj}^1} L\left(\hat{y}_{lj}^1, y_i\right) \quad \text{and} \quad E_{lj}^2 = \sum_{i \text{ with } \mathbf{x}_i \in R_{lj}^2} L\left(\hat{y}_{lj}^2, y_i\right)$$

The reduction in total error provided by the split on x_j at $node_{lj}$ is thus

$$D_{lj} = E_{lj} - \left(E_{lj}^1 + E_{lj}^2\right)$$

(In regression/SEL contexts, this is a reduction in error sum of squares provided by the split at $node_{lj}$. In 0-1 loss classification contexts it is a reduction in training set misclassification errors.)

Predictor variable importance

One might then take

$$I_j = \sum_{l=1}^{m(j)} D_{lj}$$

as a measure of the importance of x_j in fitting the tree and compare the various I_js (*or perhaps the square roots*, $\sqrt{I_j}$s).

Further, if a predictor is a (weighted) sum of regression trees (e.g. produced by "boosting" or in a "random forest") and I_{jm} measures the importance of x_j in the mth tree, then

$$I_{j.} = \frac{1}{M} \sum_{m=1}^{M} I_{jm}$$

is one measure of the importance of x_j in the overall predictor. One can then compare the various $I_{j.}$ (or square roots) as a means of comparing the importance of the input variables.